



(12) **DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION**

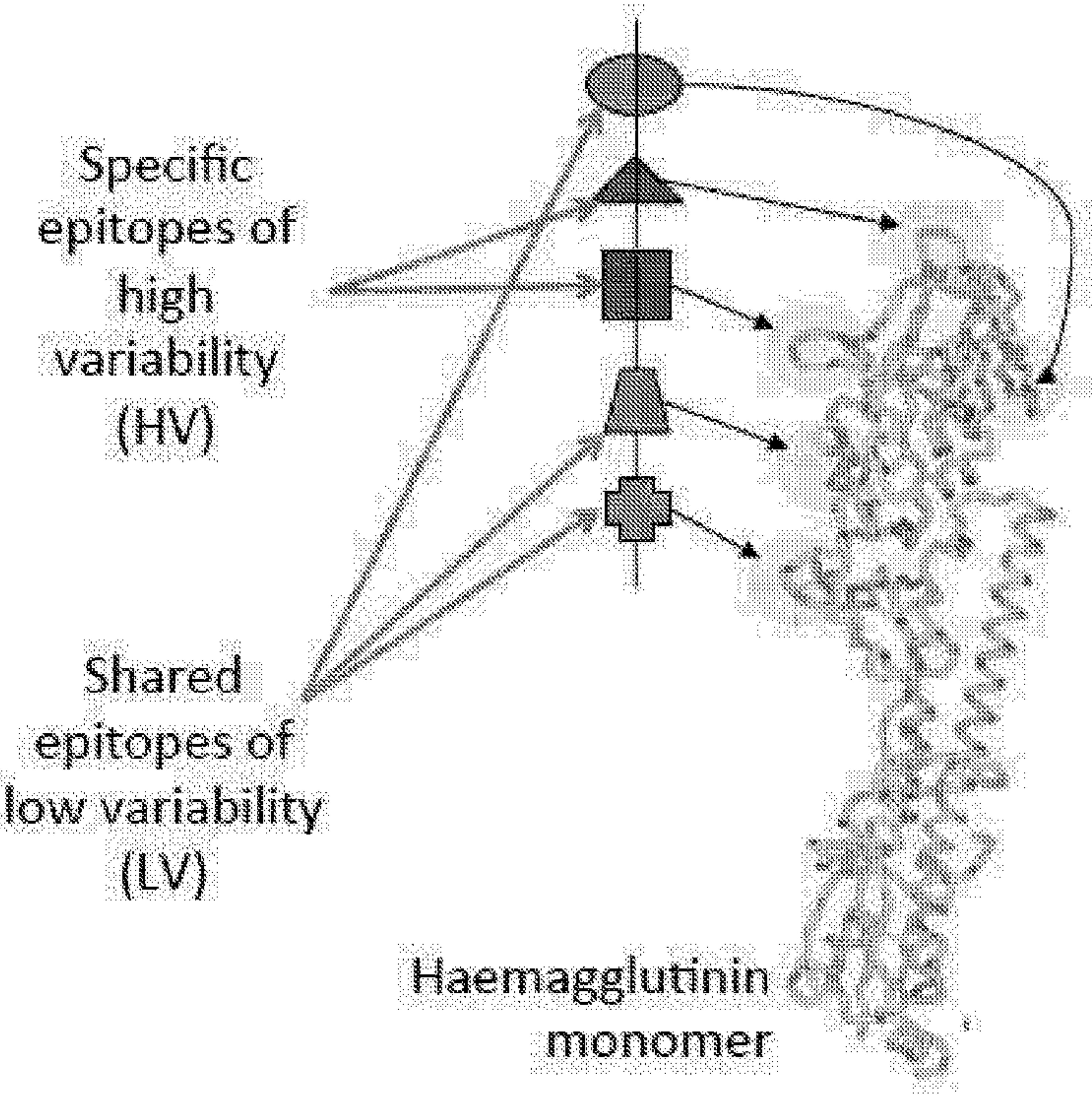
(13) **A1**

(86) Date de dépôt PCT/PCT Filing Date: 2017/08/25
(87) Date publication PCT/PCT Publication Date: 2018/03/01
(85) Entrée phase nationale/National Entry: 2019/02/19
(86) N° demande PCT/PCT Application No.: GB 2017/052510
(87) N° publication PCT/PCT Publication No.: 2018/037246
(30) Priorité/Priority: 2016/08/25 (GB1614485.9)

(51) Cl.Int./Int.Cl. *C07K 16/08* (2006.01),
C07K 16/10 (2006.01)
(71) Demandeur/Applicant:
OXFORD UNIVERSITY INNOVATION LIMITED, GB
(72) Inventeurs/Inventors:
THOMPSON, CRAIG, GB;
GUPTA, SUNETRA, GB
(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : COMPOSITION IMMUNOGENE
(54) Title: IMMUNOGENIC COMPOSITION

Figure 1



(57) **Abrégé/Abstract:**
The present invention relates to an immunogenic composition comprising two or more polypeptides. The invention also provides nucleic acid molecules and vectors encoding the polypeptides, and methods of using the compositions, nucleic acid molecules and vectors for the prevention or treatment of influenza.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
01 March 2018 (01.03.2018)



(10) International Publication Number
WO 2018/037246 A4

- (51) **International Patent Classification:**
C07K 16/08 (2006.01) *C07K 16/10* (2006.01)
- (21) **International Application Number:**
PCT/GB2017/052510
- (22) **International Filing Date:**
25 August 2017 (25.08.2017)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1614485.9 25 August 2016 (25.08.2016) GB
- (71) **Applicant: OXFORD UNIVERSITY INNOVATION LIMITED** [GB/GB]; Buxton Court, 3 West Way, Oxford OX2 0JB (GB).
- (72) **Inventors: THOMPSON, Craig**; Department of Zoology, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road Oxford Oxfordshire OX1 3SY (GB). **GUPTA, Sunetra**; Department of Zoology, University of Oxford, Peter Medawar Building for Pathogen Research, South Parks Road Oxford Oxfordshire OX1 3SY (GB).

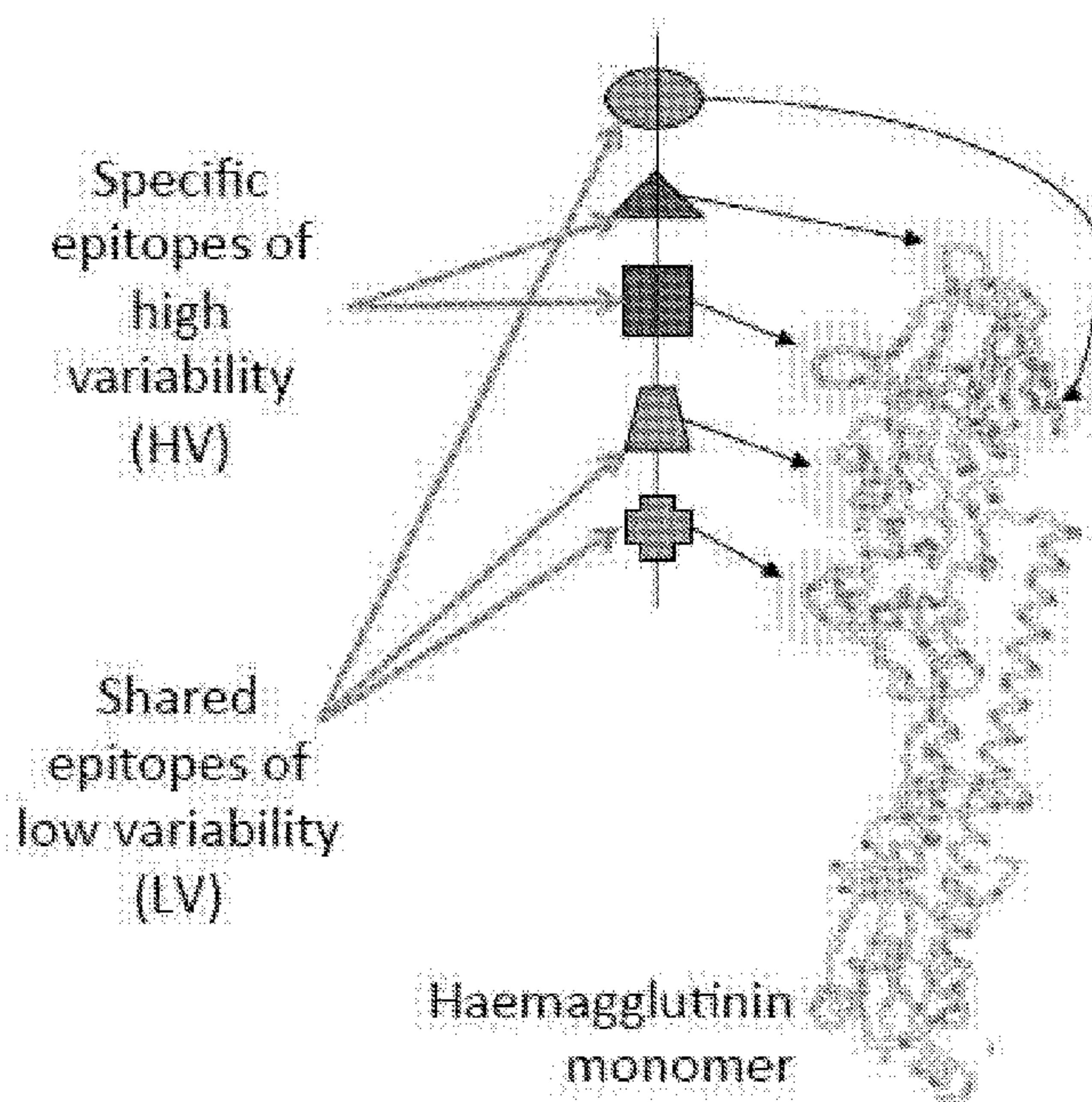
(74) **Agent: DEHNS**; St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

(81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title: IMMUNOGENIC COMPOSITION**

Figure 1



(57) **Abstract:** The present invention relates to an immunogenic composition comprising two or more polypeptides. The invention also provides nucleic acid molecules and vectors encoding the polypeptides, and methods of using the compositions, nucleic acid molecules and vectors for the prevention or treatment of influenza.

[Continued on next page]



WO 2018/037246 A4

WO 2018/037246 A4



Published:

- *with international search report (Art. 21(3))*
- *with amended claims and statement (Art. 19(1))*
- *with sequence listing part of description (Rule 5.2(a))*

Date of publication of the amended claims and statement:

19 April 2018 (19.04.2018)

IMMUNOGENIC COMPOSITION

The present invention relates to an immunogenic composition comprising two or more
5 polypeptides. The invention also provides nucleic acid molecules and vectors encoding
the polypeptides, and methods of using the compositions, nucleic acid molecules and
vectors for the prevention or treatment of influenza.

Seasonal influenza is a serious public health problem that causes severe illness and
10 death. Worldwide, seasonal influenza is estimated to cause 3 to 5 million cases of
severe illness and 250,000 to 500,000 deaths (Lozano *et al.* 2012). The demographics
highest at risk of complications are children younger than 2 years of age, adults aged
over 65, pregnant women, and people of any age with certain medical conditions such
as diabetes or weakened immune systems (Mertz *et al.* 2013). It is estimated that a
15 large proportion of child deaths in developing countries are associated with influenza.
Seasonal influenza also causes high levels of workforce absenteeism and productivity
losses.

Influenza pandemics occur sporadically when a distinct influenza strain from an animal
20 reservoir begins to circulate widely in the human population. The most recent influenza
pandemic occurred in 2009, which caused an increase in severe influenza illness and
hospitalisation in individuals aged under 35 (Presonis *et al.*, 2011; Manicassamy *et al.*,
2010). The 1918 influenza pandemic was the most serious pandemic in recorded
history, causing 50-100 million deaths. The emergence of a new pandemic influenza
25 strain remains of concern.

The most effective way to prevent illness from influenza infection is vaccination.
Currently, vaccination against influenza involves a trivalent or quadrivalent vaccine
consisting of the most recent circulating strains of the H1N1 and H3N2 subtypes of
30 influenza A and also includes one or two of influenza B strains (WHO, 2016). Due to the
rapid antigenic evolution of influenza, the vaccine has to be constantly updated, and
often due to time constraints, the wrong vaccine strains for the coming influenza season

- 2 -

are chosen. For these reasons, the convention trivalent vaccine is estimated to have 10-60% efficacy and immunisation of at risk groups takes place annually (Treanor *et al.* 2012; Belongia *et al.* 2009).

5 Consequently, there are clear societal and economic benefits for improving the current influenza vaccines. This has been recognised by pharmaceutical companies, such as GSK and Pfizer, who are developing their own new influenza vaccines. Such approaches typically target epitopes that are under weak immune selection and therefore 'immunorecessive'.

10

The influenza virus is currently conceptualised as containing (i) highly immunogenic (and protective) epitopes of high variability, as well as (ii) invariant epitopes of low immunogenicity. Together, these form the backbone of the theory of "antigenic drift" whereby the virus population slowly and incrementally acquires changes in the highly
15 variable epitope regions requiring vaccines directed against these sites to be continuously updated, with the only other alternative being seen as the artificial boosting of immunity to invariant epitopes of low natural efficacy.

20

The inventors propose, by contrast, that the influenza virus also contains highly immunogenic epitopes of low variability and that universal vaccines may be constructed by identifying these protective epitopes. This idea is underpinned by an alternative theory of influenza evolution known as "antigenic thrift" in which viral dynamics are driven by pre-existing immunity to shared epitopes, but the existence of such epitopes has remained in doubt and their use in vaccination has never previously been mooted.

25

Using a combination of bioinformatics, structural and serological analyses, one epitope of limited variability that is under strong immune selection in the major influenza antigen, haemagglutinin (HA), has now been identified and characterised.

30

HA is the major surface antigen in influenza viruses. It binds sialic acid and initiates membrane fusion, leading to endocytosis. It is a trimeric protein typically 565/566 amino acids in length. Each monomer consists of a head domain and a stem domain.

- 3 -

The epitope which has now been identified is under strong immune selection and is therefore 'immunodominant'. This has enabled the design of a new 'universal' influenza vaccine that protects against the majority of H1N1 influenza strains by targeting this
5 epitope of limited variability.

Consequently, the vaccine of the current invention has a number of advantages over the conventional trivalent vaccine and other influenza vaccines in development.

10 These advantages include:

(i) infection with circulating influenza strains will reinforce vaccine protection instead of potentially detracting from it;

(ii) the vaccine should be more immunogenic than other 'universal' vaccines in development, leading to lower thresholds of protection and greater longevity of
15 protection;

(iii) it should only need to be administered between one and three times (i.e. a prime and a boost, or a prime and two boosts); and

(iv) the theoretical and experimental framework from which the vaccine has been derived suggests that H1N1 influenza is not likely to escape the protection conferred by
20 the proposed vaccine.

It is therefore an object of the invention to provide an influenza vaccine composition which is capable of conferring protection against one or more influenza A subtypes, preferably against the H1N1 subtype.
25

In one embodiment, therefore, the invention provides an immunogenic composition comprising two or more polypeptides, wherein each polypeptide independently comprises a first region of contiguous amino acids, wherein:

30 (a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and

- 4 -

(b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

5 position 146 is T, N, I or A

position 147 is a positively charged amino acid, I or is absent

position 148 is G

position 149 is V

position 151 is A

10 position 154 is S or P

position 155 is H

position 156 is a positively charged amino acid or A or G or N or E

position 157 is a positively charged amino acid or A or G

position 158 is a positively charged amino acid or A or S or N or C or E

15 position 159 is K or A or S or N or C

position 163 is a positively charged amino acid,

wherein the amino acid sequences of the two or more polypeptides are different, and

wherein the composition is capable of inducing antibodies in a subject against an

influenza A virus, optionally together with one or more pharmaceutically-acceptable

20 carriers, adjuvants, excipients or diluents.

The invention also provides a polypeptide, wherein the amino acid sequence of the polypeptide comprises a first region, wherein:

25 (a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A subtype H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17 or H18 haemagglutinin head domain, preferably an influenza A subtype H1, H5, H6, H9 or H11 haemagglutinin head domain;

30 and

(b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

- 5 -

position 83 is E

position 85 is a negatively charged amino acid

position 146 is T, N, I or A

position 147 is a positively charged amino acid, I or is absent

5 position 148 is G

position 149 is V

position 151 is A

position 154 is S or P

position 155 is H

10 position 156 is a positively charged amino acid or A or G or N or E

position 157 is a positively charged amino acid or A or G

position 158 is a positively charged amino acid or A or S or N or C or E

position 159 is K or A or S or N or C

position 163 is a positively charged amino acid.

15

The invention also provides a composition (preferably an immunogenic composition wherein the composition is capable of inducing antibodies in a subject against an influenza A virus) comprising such a polypeptide, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents.

20

The invention also provides nucleic acids molecules (preferably DNA molecules) coding for such polypeptides, preferably wherein the DNA molecule is a vector or plasmid.

In one preferred embodiment, (b) the first region has one or more amino acid

25 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E or D

position 146 is T or N

position 147 is R or K or I or is absent

30 position 148 is G

position 149 is V

position 151 is A

- 6 -

position 154 is S or P

position 155 is H

position 156 is A or G or K or N or E

position 157 is A or G

5 position 158 is A or K

position 159 is A, K, C, N or S

position 163 is K or R.

In another preferred embodiment, (b) the first region has one or more amino acid

10 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

position 146 is T, N, I or A

position 147 is a positively charged amino acid,

15 position 148 is G

position 149 is V

position 151 is A

position 154 is S or P

position 155 is H

20 position 156 is A

position 157 is G

position 158 is K or A or S or N or C

position 159 is K or A or S or N or C

position 163 is a positively charged amino acid.

25

In another preferred embodiment, (b) the first region has one or more amino acid

substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E

30 position 146 is N

position 147 is R

position 148 is G

- 7 -

position 149 is V

position 151 is A

position 154 is P

position 155 is H

5 position 156 is A

position 157 is G

position 158 is A

position 159 is K

position 163 is K

10

Preferably, the first region of one polypeptide comprises or consists of the amino acid sequence given in SEQ ID NO: 13.

In another preferred embodiment, (b) the first region has one or more amino acid

15 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

position 146 is N

position 147 is I

20 position 148 is G

position 149 is V

position 151 is A

position 154 is S

position 155 is H

25 position 156 is K or A

position 157 is G

position 158 is A or K

position 159 is K or S

position 163 is a positively charged amino acid.

30

In another preferred embodiment, (b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

- 8 -

position 83 is E
position 85 is E
position 146 is N
position 147 is I
position 148 is G
position 149 is V
position 151 is A
position 154 is S
position 155 is H
position 156 is A
position 157 is G
position 158 is K
position 159 is S
position 163 is K

Preferably, the first region of one polypeptide comprises or consists of the amino acid sequence given in SEQ ID NO: 14.

In another preferred embodiment, (b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
position 85 is a negatively charged amino acid
position 146 is T
position 147 is a positively charged amino acid
position 148 is G
position 149 is V
position 151 is A
position 154 is S
position 155 is H
position 156 is a positively charged amino acid or A or G
position 157 is a positively charged amino acid or A or G
position 158 is K

- 9 -

position 159 is S or C

position 163 is a positively charged amino acid.

In another preferred embodiment, (b) the first region has one or more amino acid

5 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E

position 146 is T

position 147 is R

10 position 148 is G

position 149 is V

position 151 is A

position 154 is S

position 155 is H

15 position 156 is K

position 157 is G

position 158 is K

position 159 is S

position 163 is K.

20

Preferably, the first region of one polypeptide comprises or consists of the amino acid sequence given in SEQ ID NO: 15.

In another preferred embodiment, (b) the first region has one or more amino acid

25 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

position 146 is T

position 147 is a positively charged amino acid

30 position 148 is G

position 149 is V

position 151 is A

- 10 -

position 154 is S

position 155 is H

position 156 is N

position 157 is G

5 position 158 is a positively charged amino acid

position 159 is S

position 163 is a positively charged amino acid.

In another preferred embodiment, (b) the first region has one or more amino acid

10 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E

position 146 is T

position 147 is K

15 position 148 is G

position 149 is V

position 151 is A

position 154 is S

position 155 is H

20 position 156 is N

position 157 is G

position 158 is K

position 159 is S

position 163 is R.

25

Preferably, the first region of one polypeptide comprises or consists of the amino acid sequence given in SEQ ID NO: 16.

In another preferred embodiment, (b) the first region has one or more amino acid

30 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

- 11 -

position 146 is T or N

position 147 is absent

position 148 is G

position 149 is V

5 position 151 is A

position 154 is S or P

position 155 is H

position 156 is N or E

position 157 is G

10 position 158 is K or E

position 159 is S

position 163 is a positively charged amino acid.

In another preferred embodiment, (b) the first region has one or more amino acid

15 substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E

position 146 is T

position 147 is absent

20 position 148 is G

position 149 is V

position 151 is A

position 154 is S

position 155 is H

25 position 156 is N

position 157 is G

position 158 is K

position 159 is S

position 163 is R.

30

Preferably, the first region of one polypeptide comprises or consists of the amino acid sequence given in SEQ ID NO: 17.

- 12 -

In one embodiment, the invention relates to an immunogenic composition.

As used herein, the term “immunogenic” is intended to refer to the ability to elicit a specific immune response against an influenza A subtype. This response may, for example, be when a composition of the invention is administered at an appropriate dose and in an appropriate formulation which may include/require a suitable adjuvant. A booster comprising a dose similar or less than the original dose may be required to obtain the required immunogenic response.

In particular, the immunogenic composition of the invention is capable of inducing antibodies (preferably neutralising antibodies) in a subject against influenza A virus.

Preferably, the immunogenic composition of the invention is capable of providing protection in a subject against influenza A virus.

More preferably, the immunogenic composition of the invention is capable of inducing antibodies (preferably neutralising antibodies) in a subject against the H1N1 influenza A subtype.

The capability of a composition of the invention to induce neutralising antibodies in a subject (e.g. a human subject) may be tested by purifying sera from the blood of subjects to whom the composition has been administered.

Antibodies may be measured using ELISA or a pseudotype micro-neutralisation (pMN) assay. ELISA is the most sensitive of these two assays; it quantifies all antibodies. In contrast, the pMN is less sensitive, but it quantifies neutralising antibodies.

As used herein, the reference to “influenza” relates to influenza virus, preferably influenza A virus, more preferably influenza A virus H1 subtypes, and most preferably influenza A virus H1N1 subtypes.

- 13 -

The immunogenic composition comprises one, two or more polypeptides. The amino acid sequences of these two or more polypeptides are preferably different.

5 The composition may, for example, comprise 2, 3, 4, 5, 6, 7, 8, 9 or 10 different polypeptides.

Preferably, the composition comprises 2, 3, 4 or 5 different polypeptides, more preferably 3 different polypeptides.

10 The sequences of the invention are derived from or based upon the head domain of influenza A haemagglutinin proteins.

The naturally-occurring haemagglutinin protein is a homo-trimer of three polypeptides.

15 In a preferred composition of the invention, the composition comprises three polypeptides as defined herein which form a homotrimer. The composition may comprise more than one (e.g. 2, 3, 4, or 5) different homotrimers of the polypeptides defined herein.

20 In another preferred embodiment, three polypeptides of the invention form a heterotrimer in the composition. The composition may comprise more than one (e.g. 2, 3, 4, or 5) different heterotrimers of the polypeptides defined herein.

Each polypeptide independently comprises a first region of contiguous amino acids.

25 This region is a contiguous stretch of amino acids which are covalently joined.

In one embodiment, the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin (HA) head domain.

30 The intention is that this first region adopts the conformation of an influenza A haemagglutinin head domain.

- 14 -

The haemagglutinin head domain may, for example, be from any influenza A subtype, e.g. H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17 or H18.

- 5 Preferably, the haemagglutinin head domain is from an influenza A H1, H5, H6, H9 or H11 subtype.

In one embodiment, haemagglutinin head domain is from an influenza A H1 subtype.

In one embodiment, haemagglutinin head domain is from an influenza A H5 subtype.

- 10 In one embodiment, haemagglutinin head domain is from an influenza A H6 subtype.

In one embodiment, haemagglutinin head domain is from an influenza A H9 subtype.

In one embodiment, haemagglutinin head domain is from an influenza A H11 subtype.

Preferably, the influenza A N subtype is N1.

15

Consensus amino acid sequences of the H1, H5, H6 and H11 haemagglutinin polypeptides are given herein as SEQ ID NOs: 9-12, respectively. A consensus amino acid sequence of the H9 haemagglutinin polypeptide is given herein as SEQ ID NO: 23.

The H9 sequence may be used herein in place of the H1, H5, H6 or H9 embodiments

- 20 disclosed herein, *mutatis mutandis*.

The amino acid sequence number used herein is based on the numbering given to the influenza A H1 haemagglutinin head domain as given in SEQ ID NO: 9.

- 25 It should be noted that there are three ways of numbering the influenza A H1 haemagglutinin polypeptide and throughout the current specification the linear numbering is used where Met = 1.

- 30 The HA polypeptide comprises two regions: the HA1 region and the HA2 region. These regions are separated by a potential cleavage site.

- 15 -

The H1 cleavage site consensus sequence is PSIQSR/GLF (SEQ ID NO: 24); the H5 cleavage site consensus sequence is PQRKKR/GLF (SEQ ID NO: 25); the H6 cleavage site consensus sequence is PQIETR/GLF (SEQ ID NO: 26); the H9 cleavage consensus sequence is PSRSSR/GLF (SEQ ID NO: 27); and the H11 cleavage site
5 consensus sequence is PAIATR/GLF (SEQ ID NO: 28).

Cleavage of HA0 into HA1 and HA2 occurs between R/GLF and is performed by a protease. The cleavage sites stated above are all described as monobasic. In some H5 viruses, a polybasic cleavage site is present, and this differs from the monobasic sites
10 by having has multiple arginine residues (R's) and/or lysine residues (K's) in the critical position basic position.

Further details of the cleavage sites may be found in Sun *et al.*, Journal of Virology, Sept. 2010, Vol. 84, No. 17, p. 8683–8690.

15

The HA1 region comprises 1-60 amino acids of the stalk, followed by the head domain, and then additional stalk amino acids. The HA2 region comprises only stalk amino acids.

20 The head domain of haemagglutinin is defined as being between two cysteines within the HA1 region. The first cysteine is generally at position 58, 59 or 60; the second cysteine is generally at position 290, 291 or 292.

In influenza A H1 haemagglutinins, these cysteines are at positions 59 and 291/292 due
25 to the absence of an amino acid at position 147 in some H1 haemagglutinins.

For H6 haemagglutinins, the head domain corresponds to the sequence between positions 58 and 292.

30 However, for H5 and H11 haemagglutinins, the head domain corresponds to the region between positions 58 to 290, respectively.

- 16 -

For H9 haemagglutinins, the head domain corresponds to the region between positions 60 to 290.

Consensus amino acid sequences of the H1, H5, H6 and H11 haemagglutinin head
5 domains are given herein as SEQ ID NOs: 1-4, respectively.

In some embodiments, the amino acid sequence of the first region has at least 80%, 85%, 90% or 95% sequence identity to SEQ ID NO: 1.

10 In some embodiments, the amino acid sequence of the first region has at least 80%, 85%, 90% or 95% sequence identity to SEQ ID NO: 2.

In some embodiments, the amino acid sequence of the first region has at least 80%, 85%, 90% or 95% sequence identity to SEQ ID NO: 3.

15

In some embodiments, the amino acid sequence of the first region has at least 80%, 85%, 90% or 95% sequence identity SEQ ID NO: 4.

In some embodiments, the amino acid sequence of the first region has at least 65%,
20 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an influenza A haemagglutinin head domain at positions other than those that correspond to positions 83, 85, 146-149, 151, 154-159 and 163 of SEQ ID NO: 9.

In some embodiments, the amino acid sequence of the first region has at least 65%,
25 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an influenza A H1 haemagglutinin head domain (preferably of SEQ ID NO: 1) at positions other than those that correspond to positions 83, 85, 146-149, 151, 154-159 and 163 of SEQ ID NO: 9.

30 In some embodiments, the amino acid sequence of the first region has at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an influenza A H6 haemagglutinin head domain (preferably of SEQ ID NO: 2) at positions other than

- 17 -

those that correspond to positions 83, 85, 146-149, 151, 154-159 and 163 of SEQ ID NO: 9.

In some embodiments, the amino acid sequence of the first region has at least 65%,
 5 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an influenza A H5 haemagglutinin head domain (preferably of SEQ ID NO: 3) at positions other than those that correspond to positions 83, 85, 146-149, 151, 154-159 and 163 of SEQ ID NO: 9.

10 In some embodiments, the amino acid sequence of the first region has at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or 100% sequence identity to an influenza A H11 haemagglutinin head domain (preferably of SEQ ID NO: 4) at positions other than those that correspond to positions 83, 85, 146-149, 151, 154-159 and 163 of SEQ ID NO: 9.

15

In some embodiments, each polypeptide independently additionally comprises one or more amino acids which are contiguously joined to the first region at the N- and/or C-termini.

20 The additional N-terminal amino acids are preferably a stretch of contiguous amino acids which are derived from a haemagglutinin N-terminal stalk region, preferably from a haemagglutinin N-terminal stalk region of an influenza A H subtype, most preferably a H1, H5, H6, H9 or H11 subtype.

25 Preferably, 58-60 amino acids of a haemagglutinin N-terminal stalk region of an influenza A subtype are contiguously joined to the N-terminal of the first region in one or more of the polypeptides.

In some embodiments, the amino acid sequence of this stalk region has at least 65%,
 30 70%, 75%, 80%, 85%, 90%, 95% or 100% sequence identity to:

- (i) amino acids 1-59 of SEQ ID NO: 9,
- (ii) amino acids 1-58 of SEQ ID NO: 10,

- 18 -

(iii) amino acids 1-58 of SEQ ID NO: 11, or

(iv) amino acids 1-58 of SEQ ID NO: 12.

5 The additional C-terminal amino acids, if present, are preferably a stretch of contiguous amino acids (e.g. 1-300, 1-200, 1-100, 1-50 or 1-10 amino acids) which are derived from the haemagglutinin C-terminal stalk region of an influenza A subtype, preferably H1, H5 H6, H9 or H11.

10 Preferably, the stretch of contiguous amino acids is derived from the haemagglutinin stalk region of the same influenza A subtype from which the head region is derived.

In some embodiments, the polypeptides do not comprise an influenza A subtype HA2 region.

15 In some embodiments, the polypeptides do not comprise the HA2 region of SEQ ID NOs: 9-12.

20 In some preferred embodiments, one or more or all of the one, two or more polypeptides independently comprise an influenza A HA1 domain comprising a first region as defined herein, most preferably an influenza A H1, H5, H6 or H11 subtype HA1 domain comprising a first region as defined herein.

25 Preferably, the polypeptides are independently less than 600, more preferably less than 400 and most preferably less than 300 amino acids in length.

Preferably, the polypeptides are independently 250-350, more preferably 280-300 amino acids in length, and most preferably 290-292 amino acids in length.

30 The first region of the polypeptide has one or more amino acid substitutions at specified positions which correspond to positions in SEQ ID NO: 1.

- 19 -

For example, the first region of the polypeptide may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 or 14 of the specified amino acid substitutions.

5 Preferably, the first region of the polypeptide has all of the specified 14 amino acid substitutions.

As used herein, the term “positively charged amino acid” includes lysine, arginine and histidine. As used herein, the term “negatively charged amino acid” includes aspartic acid and glutamic acid.

10

In some preferred embodiments, the amino acid sequences of the polypeptides independently comprise or consist of an amino acid sequence of SEQ ID NOs: 13-17.

The polypeptides of the invention may be produced using recombinant methodology.

15 For example, such techniques are described in “Molecular Cloning: A Laboratory Manual” (Fourth Edition) Michael R. Green and Joseph Sambrook.

Alternatively, the nucleotide sequence encoding the polypeptides may be produced by chemical synthesis. Such a nucleotide sequence may then be ligated into an
20 appropriate vector for host cell transformation or transfection. The polypeptides may then be expressed in such host cells.

For modifications of existing HA genes, CRISPR-based techniques may also be used, such as those described in “CRISPR-Cas: A Laboratory Manual” (2016), edited by
25 Jennifer Doudna (University of California, Berkeley) and Prashant Mali (University of California, San Diego). TALENs-based techniques may also be used.

Alternatively, the polypeptides of the invention may be synthesised using standard chemical peptide synthesis techniques. Solid phase synthesis of peptides in which the
30 C-terminal amino acid of the sequence is attached to an insoluble support followed by sequential addition of the remaining amino acids may, for example, be used.

- 20 -

In a further embodiment, the invention provides a nucleic acid molecule which codes for one or more polypeptides of the invention. Preferably, the nucleic acid molecule encodes one, two or more, preferably 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 of the polypeptides of the invention.

5

Preferred nucleotide sequences include those comprising SEQ ID NOs: 18-22, and nucleotide sequences having at least 80%, 85%, 90% or 95% sequence identity thereto, encoding polypeptides which are capable of inducing antibodies in a subject against an influenza A virus.

10

Also preferred are nucleic acid molecules encoding polypeptide of SEQ ID NOs: 13-17.

As used herein, the terms "nucleic acid sequence", "nucleic acid molecule" and "polynucleotide" are used interchangeably and do not imply any length restriction.

15

These include DNA (including cDNA) and RNA sequences.

The nucleic acid molecules of the present invention include isolated nucleic acid molecules that have been removed from their naturally-occurring environment, recombinant or cloned DNA isolates, and chemically-synthesized analogues or analogues which have been synthesized biologically by heterologous systems.

20

The nucleic acid molecules of the present invention may be prepared by any means known in the art. For example, large amounts of the polynucleotides may be produced by replication in a suitable host cell. The natural or synthetic DNA fragments coding for a desired fragment may be incorporated into recombinant nucleic acid constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the DNA constructs will be suitable for autonomous replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to and integration within the genome of a cultured insect, mammalian, plant or other eukaryotic cell lines.

25
30

- 21 -

The nucleic acid molecules of the present invention may also be produced by chemical synthesis, e.g. by the phosphoramidite method or the tri-ester method, and may be performed on commercial automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single stranded product of chemical synthesis either
5 by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

The original (e.g. wild-type) codons in a nucleic acid molecule may be optimised for
10 expression in a desired cell line, for example, using an online tool such as that available at <http://genomes.urv.es/OPTIMIZER/>.

In one embodiment of the invention, therefore, the nucleic acid molecule is codon-optimized for expression in a host cell, preferably a human cell.

15

As used herein, the term “product of the invention” refers to the polypeptides of the invention, nucleic acids of the invention, vectors of the invention, particles of the invention and compositions of the invention, *inter alia*.

20 The invention also provides a vector or plasmid comprising a nucleic acid molecule of the invention. Preferably, the vector is an expression vector.

The vector and/or plasmid may comprise one or more regulatory sequences which are operably linked to the sequence which encodes the polypeptide, e.g. one or more
25 enhancer, promoter and/or transcriptional terminator sequences.

In a particularly preferred embodiment, there is provided an immunogenic composition comprising one, two or more vectors encoding polypeptides of SEQ ID NOs: 13-17, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants,
30 excipients or diluents, as a combined preparation in a form suitable for simultaneous, separate or sequential use for treating or preventing influenza A infection.

- 22 -

Preferably, a prime is administered to the subject with vector(s) encoding SEQ ID NOs: 14 and 15; then a first boost using vector(s) encoding SEQ ID NOs: 13 and 16; and then a final boost using a vector encoding SEQ ID NO: 17.

5 In some embodiments, the vector is viral vector, e.g. a poxvirus vector.

In other embodiments, the vector is an adenoviral vector or a Modified Vaccinia Ankara (MVA) viral vector.

10 Preferably, the vector is a non-replicating vector.

Non-replicating poxviruses and adenoviruses represent groups of viruses which may be used as vectors for the delivery of genetic material into a target cell. Viral vectors serve as antigen delivery vehicles and also have the power to activate the innate immune
15 system through binding cell surface molecules that recognise viral elements. A recombinant viral vector can be produced that carries nucleic acid encoding a given antigen. The viral vector can then be used to deliver the nucleic acid to a target cell, where the encoded antigen is produced by the target cell's own molecular machinery. As "non-self", the produced antigen generates an immune response in the target subject.

20

Without wishing to be bound by any one particular theory, the inventors believe that antigen delivery using the vectors of the invention stimulates, amongst other responses, a T-cell response in the subject. Thus, the inventors believe that one way in which the present invention provides for protection against influenza infection is by stimulating T-
25 cell responses and the cell-mediated immunity system. In addition, humoral (antibody) based protection can also be achieved.

The vector of the invention may be a non-replicating poxvirus vector. As used herein, a non-replicating (or replication-deficient) viral vector is a viral vector which lacks the
30 ability to productively replicate following infection of a target cell. Thus, a non-replicating viral vector cannot produce copies of itself following infection of a target cell. Non-

- 23 -

replicating viral vectors may therefore advantageously have an improved safety profile as compared to replication-competent viral vectors.

In one embodiment, the non-replicating poxvirus vector is selected from a Modified
5 Vaccinia virus Ankara (MVA) vector, a NYVAC vaccinia virus vector, a canarypox (ALVAC) vector, and a fowlpox (FPV) vector. MVA and NYVAC are both attenuated derivatives of vaccinia virus. Compared to vaccinia virus, MVA lacks approximately 26 of the approximately 200 open reading frames.

10 In one embodiment, the non-replicating poxvirus vector is an MVA vector.

The vector of the invention may be an adenovirus vector. In one embodiment, the adenovirus vector is a non-replicating adenovirus vector (wherein non-replicating is defined as above). Adenoviruses can be rendered non-replicating by deletion of the EI
15 or both the EI and E3 gene regions. Alternatively, an adenovirus may be rendered non-replicating by alteration of the EI or of the EI and E3 gene regions such that said gene regions are rendered non-functional. For example, a non-replicating adenovirus may lack a functional EI region or may lack functional EI and E3 gene regions. In this way the adenoviruses are rendered replication-incompetent in most mammalian cell lines and do
20 not replicate in immunised mammals. Most preferably, both EI and E3 gene region deletions are present in the adenovirus, thus allowing a greater size of transgene to be inserted. This is particularly important to allow larger antigens to be expressed, or when multiple antigens are to be expressed in a single vector, or when a large promoter sequence, such as the CMV promoter, is used. Deletion of the E3 as well as the EI
25 region is particularly favoured for recombinant Ad5 vectors. Optionally, the E4 region can also be engineered.

In one embodiment, the adenovirus vector is selected from a human adenovirus vector, a simian adenovirus vector, a group B adenovirus vector, a group C adenovirus vector,
30 a group E adenovirus vector, an adenovirus 6 vector, a PanAd3 vector, an adenovirus C3 vector, a ChAdY25 vector, an AdC68 vector, and an Ad5 vector.

- 24 -

The viral vector of the invention, as described above, can be used to deliver a single antigen to a target cell. Advantageously, the viral vector of the invention can also be used to deliver multiple (different) antigens to a target cell.

- 5 In one embodiment, the vector of the invention further comprises a nucleic acid sequence encoding an adjuvant (for example, a cholera toxin, an *E. coli* lethal toxin, or a flagellin).

10 The nucleic acid sequence encoding a vector (as described above) may be generated by the use of any technique for manipulating and generating recombinant nucleic acid known in the art. In one aspect, the invention provides a method of making a vector (as described above), comprising providing a nucleic acid, wherein the nucleic acid comprises a nucleic acid molecule encoding a vector of the invention; transfecting a host cell with the nucleic acid molecule; culturing the host cell under conditions suitable
15 for the propagation of the vector; and obtaining the vector from the host cell.

As used herein, "transfecting" may mean any non-viral method of introducing nucleic acid molecules into a cell. The nucleic acid molecule may be any nucleic acid molecule suitable for transfecting a host cell. Thus, in one embodiment, the nucleic acid molecule
20 is a plasmid. The host cell may be any cell in which a vector (i.e. a non-replicating poxvirus vector or an adenovirus vector, as described above) may be grown. As used herein, "culturing the host cell under conditions suitable for the propagation of the vector" means using any cell culture conditions and techniques known in the art which are suitable for the chosen host cell, and which enable the vector to be produced in the
25 host cell. As used herein, "obtaining the vector", means using any technique known in the art that is suitable for separating the vector from the host cell. Thus, the host cells may be lysed to release the vector. The vector may subsequently be isolated and purified using any suitable method or methods known in the art.

- 30 The invention also provides a host cell comprising a nucleic acid molecule, vector or plasmid of the invention. Preferably, the host cell is a eukaryotic host cell. Examples of eukaryotic host cells include yeast and mammalian cells.

- 25 -

The host cell is preferably a cell in which a vector (e.g. a non-replicating poxvirus vector or an adenovirus vector, as described above) may be grown or propagated. The host cell may be selected from a 293 cell (also known as a HEK, or human embryonic kidney, cell), a CHO cell (Chinese Hamster Ovary), a CCL81.1 cell, a Vero cell, a HELA cell, a
5 Per.C6 cell, a BHK cell (Baby Hamster Kidney), a primary CEF cell (Chicken Embryo Fibroblast), a duck embryo fibroblast cell, or a DF-1 cell.

In other embodiments, the host cell is a human cell (e.g. an isolated human cell).

10 In a further embodiment, there is provided a virus-like particle (VLP) comprising one, two or more (e.g. 3, 4, 5, 6, 7, 8, 9 or 10) polypeptides of the invention. The particle is preferably immunogenic.

Virus-like particles resemble viruses, but are non-infectious because they do not contain
15 any viral genetic material. The particles may also be described as multimeric lipoprotein particles.

Once expressed in an appropriate system, these VLPs are able to assemble spontaneously into lipoprotein structures/particles composed of one or more monomers
20 of said polypeptides.

The invention also provides a VLP wherein one, two or more (e.g. 3, 4, 5, 6, 7, 8, 9 or 10) polypeptides (preferably different polypeptides) of the invention are covalently attached to the VLP. For example, the polypeptides of the invention may be covalently
25 attached to the VLP by using chemical cross-linkers, reactive unnatural amino acids or SpyTag/SpyCatcher reactions.

In a particularly preferred embodiment, there is provided an immunogenic composition comprising at least five different virus-like particles (VLPs), wherein each VLP
30 independently comprises one or more homotrimers consisting of or comprised of polypeptides of SEQ ID NOs: 13-17, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents, as a combined

- 26 -

preparation in a form suitable for simultaneous, separate or sequential use for treating or preventing influenza A infection.

Preferably, a prime is administered to the subject with homotrimers of SEQ ID NOs: 14 and 15; the first boost using homotrimers of SEQ ID NOs: 13 and 16; and the final boost using a homotrimer of SEQ ID NO: 17.

The invention also provides a composition comprising one, two or more polypeptides of the invention, one or more nucleic acid molecules of the invention, one or more vectors of the invention or a VLP of the invention, optionally together with one or more pharmaceutically-acceptable carriers, excipients or diluents.

The composition is preferably an immunogenic composition.

Substances suitable for use as pharmaceutically-acceptable carriers are known in the art. Non-limiting examples of pharmaceutically-acceptable carriers include water, saline, and phosphate-buffered saline. In some embodiments, however, the composition is in lyophilized form, in which case it may include a stabilizer, such as bovine serum albumin (BSA). In some embodiments, it may be desirable to formulate the composition with a preservative, such as thiomersal or sodium azide, to facilitate long term storage. Examples of buffering agents include, but are not limited to, sodium succinate (pH 6.5), and phosphate buffered saline (PBS; pH 7.4).

In addition to a pharmaceutically-acceptable carrier, the composition of the invention can be further combined with one or more of a salt, excipient, diluent, adjuvant, immunoregulatory agent and/or antimicrobial compound.

In one embodiment, the products of the invention may contain 5% to 95% of active ingredient (i.e. polypeptide, nucleic acid, vectors or VLPs), such as at least 10% or 25% of active ingredient, or at least 40% of active ingredient or at least 50%, 55%, 60%, 70% or 75% active ingredient.

- 27 -

The products of the invention may be administered in a manner compatible with the dosage formulation, and in such amount as will be prophylactically and/or therapeutically effective.

- 5 Administration of the products of the invention is generally by conventional routes, e.g. intravenous, subcutaneous, intraperitoneal, or mucosal routes. The administration may be by parenteral administration; for example, a subcutaneous or intramuscular injection.

Accordingly, the products of the invention may be prepared as injectables, either as
10 liquid solutions or suspensions. Solid forms suitable for solution in, or suspension in, liquid prior to injection may alternatively be prepared. The preparation may also be emulsified, or the peptide encapsulated in liposomes or microcapsules. The active ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline,
15 dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the products of the invention may also contain minor amounts of auxiliary substances such as wetting or emulsifying agents, and/or pH buffering agents.

Additional formulations which are suitable for other modes of administration include oral
20 formulations or formulations suitable for distribution as aerosols. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders.

25

It may be desired to direct the products of the present invention (as described above) to the respiratory system of a subject. Efficient transmission of a therapeutic/prophylactic composition or medicament to the site of infection in the lungs may be achieved by oral or intra-nasal administration.

30

Formulations for intranasal administration may be in the form of nasal droplets or a nasal spray. An intranasal formulation may comprise droplets having approximate

diameters in the range of 100-5000 μm , such as 500-4000 μm , 1000-3000 μm or 100-1000 μm . Alternatively, in terms of volume, the droplets may be in the range of about 0.001-100 μl , such as 0.1-50 μl or 1.0-25 μl , or such as 0.001-1 μl .

5 Alternatively, the therapeutic/prophylactic formulation or medicament may be an aerosol formulation. The aerosol formulation may take the form of a powder, suspension or solution. The size of aerosol particles is relevant to the delivery capability of an aerosol. Smaller particles may travel further down the respiratory airway towards the alveoli than would larger particles. In one embodiment, the aerosol particles have a diameter
10 distribution to facilitate delivery along the entire length of the bronchi, bronchioles, and alveoli. Alternatively, the particle size distribution may be selected to target a particular section of the respiratory airway, for example the alveoli. In the case of aerosol delivery of the medicament, the particles may have diameters in the approximate range of 0.1-50 μm , preferably 1-25 μm , more preferably 1-5 μm .

15

Aerosol particles may be for delivery using a nebulizer (e.g. via the mouth) or nasal spray. An aerosol formulation may optionally contain a propellant and/or surfactant.

20

Preferably, the composition of the invention is a vaccine composition, e.g. suitable for parenteral administration, optionally together with one or more adjuvants.

25

As used herein, a vaccine is a formulation that, when administered to an animal subject such as a mammal (e.g. a human, bovine, porcine, ovine, caprine, equine, cervine, canine or feline subject; in particular a human subject), stimulates a protective immune response against an infectious disease. The immune response may be a humoral and/or a cell-mediated immune response. Thus, the vaccine may stimulate B cells and/or T cells.

30

Examples of suitable adjuvants include those which are selected from the group consisting of:

- metal salts such as aluminium hydroxide or aluminium phosphate,
- oil in water emulsions,

- 29 -

- toll like receptors agonist, (such as toll like receptor 2 agonist, toll like receptor 3 agonist, toll like receptor 4 agonist, toll like receptor 7 agonist, toll like receptor 8 agonist and toll like receptor 9 agonist),
- saponins, for example Quil A and its derivatives such as QS7 and/or QS21,
- 5 - CpG containing oligonucleotides,
- 3D -MPL,
- (2-deoxy-6-o-[2-deoxy-2-[(R)-3-dodecanoyloxytetradecanoylamino]-4-o-phosphono- β -D-glucopyranosyl]-2-[(R)-3-hydroxytetradecanoylamino]- α -D-glucopyranosyldihydrogenphosphate),
- 10 - DP (3S, 9 R) -3-[(R)-dodecanoyloxytetradecanoylamino]-4-oxo-5-aza-9(R)-[(R)-3 -hydroxytetradecanoylamino] decan- 1, 10-diol, 1,10- bis(dihydrogenophosphate), and
- MP-Ac DP (3S-, 9R) -3-[(R) -dodecanoyloxytetradecanoylamino]-4-oxo-5- aza-9-[(R)-3-hydroxytetradecanoylamino]decan-I , 10-diol, 1-dihydrogenophosphate 10-(6-
- 15 aminohexanoate),
- or combinations thereof.

Preferably, the adjuvant is selected from the group comprising:

- a saponin associated with a metallic salt, such as aluminium hydroxide or aluminium
- 20 phosphate
- 3D-MPL, QS21 and a CpG oligonucleotide, for example as an oil in water formulation,
- saponin in the form of a liposome, for example further comprise a sterol such as QS21 and sterol, and
- ISCOM.

25

In some particularly preferred embodiments, the adjuvant comprises a saponin.

Saponins are steroid or triterpenoid glycosides, which occur in many plant species.

Saponin-based adjuvants act in part by stimulating the entry of antigen-presenting cells into the injection site and enhancing antigen presentation in the local lymph nodes.

30

Preferably, the adjuvant comprises saponin, cholesterol and a phospholipid, e.g. ISCOM Matrix-M™ (Isconova, Novavax).

- 30 -

In Matrix-M, purified saponin fractions are mixed with synthetic cholesterol and a phospholipid to form stable particles than can be readily formulated with a variety of vaccine antigens. Matrix-M™ induces both a cell-mediated and an antibody mediated
5 immune response.

In some other preferred embodiments, the adjuvant comprises a squalene-oil-in-water nano-emulsion emulsion, e.g. AddaVax™ (InvivoGen).

10 Squalene is an oil which is more readily metabolized than the paraffin oil used in Freund's adjuvants. Squalene oil-in-water emulsions are known to elicit both cellular (Th1) and humoral (Th2) immune responses. This class of adjuvants is believed to act through recruitment and activation of APC and stimulation of cytokines and chemokines production by macrophages and granulocytes.

15 The composition may further comprise a surfactant. Examples of suitable surfactants include Tween (such as Tween 20), brij and polyethylene glycol.

Vaccine preparation is generally described in New Trends and Developments in
20 Vaccines, edited by Voller *et al.*, University Park Press, Baltimore, Maryland, U.S.A., 1978. Encapsulation within liposomes is described, for example, by Fullerton, U.S. Patent 4,235,877.

The amount of the polypeptide, nucleic acid molecule, vector, or particle of the present
25 invention present in each vaccine dose is selected as an amount which induces an immunoprotective response without significant, adverse side effects in typical vaccines. Such amount will vary depending upon which specific immunogen is employed and whether or not the vaccine is adjuvanted. Generally, it is expected that each does will comprise 1-1000µg of protein, for example 1-200 µg, such as 10-100µg, and more
30 particularly 10-40µg. An optimal amount for a particular vaccine can be ascertained by standard studies involving observation of antibody titres and other responses in subjects. Following an initial vaccination, subjects will preferably receive a boost in about 4 weeks,

- 31 -

followed by repeated boosts every six months for as long as a risk of infection exists. The immune response to the products of this invention is enhanced by the use of adjuvant and or an immunostimulant.

- 5 The amount of saponin for use in the adjuvants of the present invention may be in the region of 1-1000µg per dose, generally 1-500µg per dose, more such as 1-250µg per dose, and more specifically between 1 to 100µg per dose (e.g. 10, 20, 30, 40, 50, 60, 70, 80 or 90µg per dose).
- 10 The invention also provides a combined preparation comprising two or more components selected from two or more polypeptides of the invention, two or more particles of the invention, two or more nucleic acids of the invention, two or more vectors of the invention and two or more compositions of the invention as a combined preparation in a form suitable for simultaneous, separate or sequential use, preferably
- 15 for treating or preventing influenza A infection.

In yet another aspect, the invention provides an antibody against a polypeptide of the invention.

- 20 In yet further embodiments, the invention provides a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention for use in therapy or for use as a medicament.

- In a further aspect, the invention provides a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention for use in a method of preventing or treating influenza infection in a subject.
- 25

- In a further aspect, the invention provides a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention for use in a method of inducing a T-cell or B-cell response to an influenza antigen in a subject.
- 30

- 32 -

In particular, a non-replicating poxvirus vector of the invention can be used to stimulate a protective immune response via the cell-mediated immune system. In one embodiment, the T-cell is a T-helper cell (T_h-cell). In one embodiment, the T-cell is a T_h17-cell.

In further embodiments, the invention provides the use of a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention in the manufacture of a medicament for use in a method of preventing or treating an influenza infection in a subject.

In further embodiments, the invention provides the use of a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention in the manufacture of a medicament for use in a method of inducing a T cell or B-cell response to an influenza antigen in a subject.

The invention also provides a method of treating a subject susceptible to influenza infection comprising administering an effective amount of a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention to the subject.

The invention also provides a method of inducing a T-cell or B-cell response to an influenza antigen in a subject comprising administering an effective amount of a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention to the subject.

A polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention may also be used in similar uses and methods to produce neutralising antibodies *in vivo* against influenza antigens.

Preferably, the influenza antigen is the haemagglutinin protein, more preferably, the HA1 or head domain of a haemagglutinin protein.

Preferably, the influenza is an influenza A.

5 The efficacy of the uses and methods to treat/prevent influenza infection may be tested (e.g. by ELISA) by establishing the presence or absence of neutralising antibodies against influenza virus in the subject's blood.

10 Also provided is an immunogenic composition comprising two or more polypeptides, two or more nucleic acid molecules or two or more vectors or plasmids as defined herein as a combined preparation in a form suitable for simultaneous, separate or sequential use for the treatment or prevention of influenza, preferably influenza A, or for inducing a T-cell or B-cell response in a subject against an influenza virus, preferably an influenza A virus.

15 The subject is preferably a mammal, more preferably a human.

As used herein, the term "preventing" includes preventing the initiation of influenza infection and/or reducing the severity of intensity of an influenza infection. Thus, "preventing" encompasses vaccination.

20

As used herein, the term "treating" embraces therapeutic and preventative/prophylactic measures (including post-exposure prophylaxis) and includes post-infection therapy and amelioration of an influenza infection. Each of the above-described methods and uses can comprise the step of administering to a subject an effective amount, such as a therapeutically effective amount, of a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention.

25

As used herein, an effective amount is a dosage or amount that is sufficient to achieve a desired biological outcome. As used herein, a therapeutically effective amount is an amount which is effective, upon single or multiple dose administration to a subject (such as a mammalian subject, in particular a human subject) for treating, preventing, curing,

30

- 34 -

delaying, reducing the severity of, ameliorating at least one symptom of a disorder or recurring disorder, or prolonging the survival of the subject beyond that expected in the absence of such treatment.

5 Accordingly, the quantity of active ingredient to be administered depends on the subject to be treated, capacity of the subject's immune system to generate a protective immune response, and the degree of protection required. Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be particular to each subject. Administration to the subject can comprise administering
10 to the subject a polypeptide of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention (i.e. a product of the invention) wherein the product of the invention is sequentially administered multiple times (for example, wherein the composition is administered two, three or four times). Thus, in one embodiment, the subject is administered a polypeptide of the
15 invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention and is then administered the same product of the invention (or a substantially similar product) again at a different time.

In one embodiment, administration to a subject comprises administering a polypeptide
20 of the invention, a particle of the invention, a nucleic acid of the invention, a vector of the invention or a composition of the invention to a subject, wherein said product of the invention is administered substantially prior to, simultaneously with, or subsequent to, another immunogenic composition.

25 The invention also extends to prime-boost regimes.

For example, priming and/or boosting may be effected using one or more products of the invention. The products may be administered to a subject sequentially, simultaneously or separately.

30

- 35 -

A preferred prime-boost strategy of the invention provides a method of preventing or treating an influenza infection in a subject or of inducing a T-cell or B-cell response to an influenza antigen in a subject, the method comprising the steps of:

- (i) simultaneously, separately or sequentially administering an effective amount of one,
 5 two, three, four, five or more different polypeptides to a subject in need thereof,
 wherein each polypeptide independently comprises a first region of contiguous amino acids, wherein:

- (a) the amino acid sequence of the first region has at least 80% sequence
 10 identity to an influenza A haemagglutinin head domain; and
 (b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:
- position 83 is E
 - position 85 is a negatively charged amino acid
 - 15 position 146 is T, N, I or A
 - position 147 is a positively charged amino acid, I or is absent
 - position 148 is G
 - position 149 is V
 - position 151 is A
 - 20 position 154 is S or P
 - position 155 is H
 - position 156 is a positively charged amino acid or A or G or N or E
 - position 157 is a positively charged amino acid or A or G
 - position 158 is a positively charged amino acid or A or S or N or C or E
 - 25 position 159 is K or A or S or N or C
 - position 163 is a positively charged amino acid.

Preferred influenza A haemagglutinin head domain sequences and first region substitutions are disclosed herein, *mutatis mutandis*.

30

- 36 -

The polypeptides may be in the form of a pharmaceutical composition, preferably a vaccine composition, optionally together with one or more pharmaceutically-acceptable carriers, diluents, excipients and adjuvants.

- 5 Preferably, one or more of the polypeptides (as defined above) are in the form of one or more trimers. In some embodiments, the trimers are homotrimers. In other embodiments, the trimers are heterotrimers.

Preferably, the method comprises the additional steps of:

- 10 (ii) administering a boost with a second polypeptide to the subject; and optionally also
(iii) administering a boost with a third polypeptide to the subject,
wherein the second and third polypeptides (as defined above) are preferably different to each other and preferably different to the first polypeptide.

15

Preferably, the method comprises the additional steps of:

- (ii) administering a boost with a second trimer to the subject; and optionally also
(iii) administering a boost with a third trimer to the subject,
wherein the second and third trimers are preferably different to each other and
20 preferably different to the first trimer.

Preferably, the first, second and third polypeptides are independently selected from the group consisting of polypeptides comprising or consisting of SEQ ID NOs: 13-17.

- 25 Preferably, the first, second and third trimers independently consist of polypeptides comprising or consisting of SEQ ID NOs: 13-17.

In a preferred embodiment, polypeptides or homotrimers of SEQ ID NOs: 14 and 15 are first administered to the subject; polypeptides or homotrimers of SEQ ID NOs: 13 and
30 16 are next administered to the subject; and polypeptides or homotrimers of SEQ ID NO: 17 are then administered to the subject.

- 37 -

In another preferred embodiment, the polypeptides or trimers are administered in the form of a VLP, i.e. a VLP is administered which comprises the polypeptide(s) trimer(s).

5 In another preferred embodiment, a nucleic acid molecule (preferably a vector) is administered to the subject, wherein the nucleic acid molecule encodes one or more of the polypeptides as defined above. Preferred vectors are discussed herein.

10 In one embodiment, the first and second products are administered as part of a prime-boost administration protocol. Thus, the first product may be administered to a subject as the "prime" and the second product subsequently administered to the same subject as the "boost".

15 In one embodiment, the first product is an adenovirus vector of the invention prime, and the second product is a non-replicating poxvirus vector of the invention boost.

In one embodiment, each of the above-described methods further comprises the step of administration to the subject of a product of the invention.

20 In one embodiment, the polypeptide of the invention is administered separately from the administration of a viral vector of the invention. Preferably the polypeptide and a viral vector are administered sequentially, in any order. Thus, in one embodiment, the viral vector ("V") and the polypeptide ("P") may be administered in the order V-P, or in the order P-V.

25 In certain embodiments, the above-described methods further comprise the administration to the subject of an adjuvant. Adjuvant may be administered with any of the products of the invention.

30 The products of the invention may be given in a single dose schedule (i.e. the full dose is given at substantially one time). Alternatively, the products of the invention may be given in a multiple dose schedule.

A multiple dose schedule is one in which a primary course of treatment (e.g. vaccination) may be with 1-6 separate doses, followed by other doses given at subsequent time intervals required to maintain and or reinforce the immune response, for example (for human subjects), at 1-4 months for a second dose, and if needed, a subsequent dose(s) after a further 1-4 months.

The dosage regimen will be determined, at least in part, by the need of the individual and be dependent upon the judgment of the practitioner (e.g. doctor or veterinarian).

Simultaneous administration means administration at (substantially) the same time.

Sequential administration of two or more products of the invention means that the products are administered at (substantially) different times, one after the other.

For example, sequential administration may encompass administration of two or more products of the invention at different times, wherein the different times are separated by a number of days (for example, 1, 2, 5, 10, 15, 20, 30, 60, 90, 100, 150 or 200 days).

For example, in one embodiment, the vaccine of the present invention may be administered as part of a 'prime-boost' vaccination regime.

In one embodiment, the products of the invention can be administered to a subject such as a mammal (e.g. a human, bovine, porcine, ovine, caprine, equine, cervine, canine or feline subject) in conjunction with (simultaneously or sequentially) one or more immunoregulatory agents selected from, for example, immunoglobulins, antibiotics, interleukins (e.g. IL- 2, IL-12), and/or cytokines (e.g. IFN- γ).

In yet further embodiments, the invention provides a process for the production of a one or more of polypeptides of the invention, which process comprises expressing one or more nucleic acid molecules coding for one, two or more of said polypeptides in a suitable host, and recovering the polypeptide product(s).

- 39 -

Preferably, the host is a human cell.

There are many established algorithms available to align two amino acid sequences. Typically, one sequence acts as a reference sequence, to which test sequences may be compared. The sequence comparison algorithm calculates the percentage sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters. Alignment of amino acid sequences for comparison may be conducted, for example, by computer-implemented algorithms (e.g. GAP, BESTFIT, FASTA or TFASTA), or BLAST and BLAST 2.0 algorithms.

Percentage amino acid sequence identities and nucleotide sequence identities may be obtained using the BLAST methods of alignment (Altschul *et al.* (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402; and <http://www.ncbi.nlm.nih.gov/BLAST>). Preferably the standard or default alignment parameters are used.

Standard protein-protein BLAST (blastp) may be used for finding similar sequences in protein databases. Like other BLAST programs, blastp is designed to find local regions of similarity. When sequence similarity spans the whole sequence, blastp will also report a global alignment, which is the preferred result for protein identification purposes. Preferably the standard or default alignment parameters are used. In some instances, the "low complexity filter" may be taken off.

BLAST protein searches may also be performed with the BLASTX program, score=50, wordlength=3. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in Altschul *et al.* (1997) Nucleic Acids Res. 25: 3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. (See Altschul *et al.* (1997) *supra*). When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs may be used.

- 40 -

With regard to nucleotide sequence comparisons, MEGABLAST, discontinuous-megablast, and blastn may be used to accomplish this goal. Preferably the standard or default alignment parameters are used. MEGABLAST is specifically designed to efficiently find long alignments between very similar sequences. Discontinuous

5 MEGABLAST may be used to find nucleotide sequences which are similar, but not identical, to the nucleic acids of the invention.

The BLAST nucleotide algorithm finds similar sequences by breaking the query into short subsequences called words. The program identifies the exact matches to the

10 query words first (word hits). The BLAST program then extends these word hits in multiple steps to generate the final gapped alignments. In some embodiments, the BLAST nucleotide searches can be performed with the BLASTN program, score=100, wordlength=12.

15 One of the important parameters governing the sensitivity of BLAST searches is the word size. The most important reason that blastn is more sensitive than MEGABLAST is that it uses a shorter default word size (11). Because of this, blastn is better than MEGABLAST at finding alignments to related nucleotide sequences from other organisms. The word size is adjustable in blastn and can be reduced from the default

20 value to a minimum of 7 to increase search sensitivity.

A more sensitive search can be achieved by using the newly-introduced discontinuous megablast page (www.ncbi.nlm.nih.gov/Web/Newsltr/FallWinter02/blastlab.html). This page uses an algorithm which is similar to that reported by Ma et al. (Bioinformatics.

25 2002 Mar; 18(3): 440-5). Rather than requiring exact word matches as seeds for alignment extension, discontinuous megablast uses non-contiguous word within a longer window of template. In coding mode, the third base wobbling is taken into consideration by focusing on finding matches at the first and second codon positions while ignoring the mismatches in the third position. Searching in discontinuous

30 MEGABLAST using the same word size is more sensitive and efficient than standard blastn using the same word size. Parameters unique for discontinuous megablast are:

word size: 11 or 12; template: 16, 18, or 21; template type: coding (0), non-coding (1), or both (2).

5 In some embodiments, the BLASTP 2.5.0+ algorithm may be used (such as that available from the NCBI) using the default parameters.

In other embodiments, a BLAST Global Alignment program may be used (such as that available from the NCBI) using a Needleman-Wunsch alignment of two protein sequences with the gap costs: Existence 11 and Extension 1.

10

The method of identifying sites of limited variability and subsequent epitopes of limited variability as disclosed herein may be applied to all influenza A subtypes. In particular, since the H3 subtype influenza A virus evolves in a similar way to the H1 subtype influenza A virus, the approach disclosed herein to identifying epitopes is particularly
15 applicable to H3 subtypes of influenza A.

In yet a further embodiment, therefore, the invention provides a method for identifying an epitope on a haemagglutinin head domain of an influenza virus of a defined subtype, the method comprising the steps of:

- 20 (i) identifying a possible antibody binding site on a haemagglutinin head domain polypeptide of an influenza virus of a defined subtype;
- (ii) identifying one or more continuous or discontinuous stretches of the head domain polypeptide within the possible antibody binding site; and
- (iii) comparing the amino acid sequences of those stretches with amino acid
25 sequences of haemagglutinin head domains from a plurality of influenza strains of the defined subtype in order to identify regions of limited amino acid sequence variability within those stretches;

thereby identifying a set of positions which have limited variability in their amino acid composition within the haemagglutinin head domain of the influenza virus of the defined
30 subtype and which form an epitope.

The epitopes which are identified in this way are under strong immune selection such that they periodically repeat through a limited number of forms throughout the evolution of the influenza virus.

- 5 The method and process may be applied to any influenza virus. Preferably, the influenza virus is an influenza A virus.

The influenza virus may be of any subtype, e.g. H1, H2, H3, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, H17 or H18. Preferably, the influenza virus is of
10 the H1 or H3 subtype.

Step (i) requires identifying a possible antibody binding site on a haemagglutinin head domain polypeptide of an influenza virus of a defined subtype.

- 15 This may be done by analysing the crystal structure of the haemagglutinin head domain polypeptide. Crystal structures of two or more head domains from influenza viruses of the defined subtype may be aligned to determine which residues are present on the surface of the polypeptide and the accessibility of those residues. Typically, the accessibility of the positions are the same in all crystal structures but when a position is
20 more accessible in one crystal structure than another crystal structure, the position is allocated as being more accessible to prevent the false identification of sites of limited variability.

In silico analysis may be used to determine how the accessibility and binding site area
25 contribute to the variability of hypothetical antibody binding sites. An antibody binding site of between 600 Å² and 1000 Å² may be used to determining the variability for accessibility parameters of amino acids with between >30% and >1% accessibility.

Step (ii) requires identifying one or more continuous or discontinuous stretches of the
30 head domain polypeptide within the possible antibody binding site. These stretches may be contacted by the antibody.

- 43 -

Once possible antibody binding sites have been identified, one or more continuous or discontinuous stretches of the head domain polypeptide within the possible antibody binding site may be identified, for example, by using Swiss-pdb viewer.

- 5 Step (iii) requires comparing the amino acid sequences of the continuous or discontinuous stretches with amino acid sequences of haemagglutinin head domains from a plurality of influenza strains of the defined subtype in order to identify regions of limited amino acid sequence variability within those stretches.
- 10 For example, the plurality of influenza strain sequences may be obtained from yearly consensus sequences of the haemagglutinin head domains of the influenza strains of the defined subtype. The yearly consensus sequences may be generated by dividing curated haemagglutinin sequences into separate datasets based on the year that the sequence was collected. The R package 'seqinr' or an alternate consensus sequence
- 15 generating program may then be used to generate consensus sequences.

As used herein, the term "limited sequence variability" refers to an amino acid or sequences of amino acids which are restricted in the number of different epitope conformations they can form.

20

In other embodiments, the term "limited sequence variability" refers to amino acid positions at which 0, 1, 2, 3 or 4 (preferably 0 or 1) different amino acids were found during the amino acid sequence comparisons.

- 25 In this way, a set of conserved amino acids within the haemagglutinin head domain of the influenza virus of the defined subtype may be identified which form an epitope.

The epitope may be one which is bound by an antibody. Preferably, the epitope is an epitope of limited variability.

30

The invention also provides a process for producing a polypeptide, the process comprising the steps of:

- 44 -

(i) using a method for identifying an epitope as defined herein to identify a set of amino acids with limited variability within a haemagglutinin head domain of an influenza virus of a defined subtype;

(ii) producing a polypeptide which comprises a first region of contiguous amino acids,
5 wherein:

(a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and

(b) the first region has one or more amino acid substitutions at positions which correspond to the positions of the amino acid positions with limited variability, and
10 wherein the substitution introduces the amino acid which is conserved at that position;

wherein the polypeptide is capable of inducing antibodies in a subject against an influenza A virus.

15 The invention also provides a process for producing an immunogenic composition, the process comprising the steps of:

(i) using a method for identifying an epitope as defined herein to identify a set of amino acid positions with limited variability within a haemagglutinin head domain of an influenza virus of a defined subtype;

20 (ii) producing a polypeptide which comprises a first region of contiguous amino acids, wherein:

(a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and

(b) the first region has one or more amino acid substitutions

25 at positions which correspond to the positions of the conserved amino acids, and wherein the substitution introduces the amino acid which is conserved at that position;

wherein the polypeptide is capable of inducing antibodies in a subject against an influenza A virus; and

30 (iii) formulating one or more of the polypeptides into an immunogenic composition, wherein the composition optionally comprises one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents.

- 45 -

The composition is capable of inducing antibodies in a subject against an influenza A virus.

- 5 With regard to influenza virus H3 subtype epitopes, the amino acid sequence of the first region preferably has at least 80%, 90% 95% or 100% sequence identity to an influenza A subtype H4, H7 H10, H14 or H15 haemagglutinin head domain.

10 The immunogenic composition is preferably a vaccine composition; it may be administered as a prime-boost-boost, e.g. using the compositions and regimes described herein.

The disclosure of each reference set forth herein is specifically incorporated herein by reference in its entirety.

15

BRIEF DESCRIPTION OF THE FIGURES

Figure 1: A multi-locus representation of epitopes on a monomer of haemagglutinin (HA). Each influenza strain is assumed to contain specific epitopes of high variability as well as epitopes of low variability shared with other strains.

20

Figure 2: Cyclical replacement of dominant antigenic types. The dynamics shown is for a 3-epitope system, each containing 3 possible variants as indicated in the cartoon.

25 Figure 3: "Heat map" of plasma from children showing cyclical cross-reactivity with a number of influenza strains. The year on the left hand side of the heat map relates to the strain from which the HA1 domain was taken. Individuals are aligned from 12 to 17 months left to right. Percentage reactivity is stated to the right hand side of the heat map.

30 Figure 4: "Heat map" of plasma from children showing cross-reactivity with a number of historical influenza strains. The year on the left hand side of the heat map relates to the strain from which the HA1 domain was taken. Individuals are aligned from

6 to 11 months left to right. Percentage reactivity is stated to the right hand side of the heat map.

Figure 5: Microneutralisation assays. The x-axis refers to the ratio of the IC₅₀s of the pseudotype viruses to produce a fold-change.

(A) Microneutralisation assay using wild-type (WT) and -147K mutant A/Solomon Islands/3/2006 pseudotyped viruses.

(B) Microneutralisation assay using wild-type (WT) and -147K mutant A/Puerto Rico/8/1934 pseudotyped viruses.

(C) Microneutralisation assay using wild-type (WT) and -147K mutant A/WSN/1933 pseudotyped viruses.

Figure 6: Sequential vaccination using chimeric HA constructs.

(A) Five groups of mice were sequentially vaccinated with the sequences outlined in (B), substituted into H6, H5 and H11 HAs. Two further groups were sequentially vaccinated with H6, H5 and H11 constructs without any sequence substituted into them and named 'grey' and 'purple'. A further two groups were mock vaccinated and named 'white' and 'black'. The first two vaccinations were administered as a 100 µg intra muscular injection of DNA, whilst the final vaccination was administered as an intra-muscular injection of 8 HI units of lentivirus displaying a chimeric HA (i.e. H11 with or without substitution) with an Alum adjuvant.

(B)-(F) Pseudotype microneutralisation assays using 0.5 µl of sera from the bleed at 21 weeks. Broad neutralising activity occurs against lentiviruses displaying H1 HAs from influenza viruses circulating in 1933, 1934, 1977, 2006 and 2009.

(G)-(J) Influenza challenge of vaccinated mice with either A/PR/8/1934 or A/California/4/2009. The graphs denote daily weight loss and percentage survival of the mice during the challenge.

Figure 7: Sites of limited variability are present in the head of H1 HA.

(A) Antibody binding sites were mapped to the A/Puerto Rico/8/1934 crystal structure and the variability within those site determined by referring to an alignment of 2,756 H1 sequences. Only parts of A/Puerto Rico/8/1934 crystal structure accessible to antibody

- 47 -

binding were considered. In A. an antibody binding site of 800 Å² was used to determining the variability for three accessibility parameters: amino acids with >30%, >10% or >1% accessibility.

(B) In B. a dataset of amino acids with >10% accessibility was used to determine the variability for three binding site sizes: 600, 800 or 1000 Å². Both approaches identified the same regions within the head of H1 HA which are of limited variability. One of these regions contains our epitope of limited variability centring of position 156/158. Linear numbering of HA is used for the x-axis.

(C)-(J) Mapping of predicted antibody-binding sites onto the crystal structures of HA domains from specified influenza strains.

Figure 8: The disrupted peptide sequence corresponding to the site surrounding amino acids 156/158. (A) shows the crystal structures of A/Brevig Mission/1/1918 and (B) shows the crystal structure of A/Puerto Rico/8/1934 from the side and above. The disrupted peptide sequence is mapped on the H1 structures. Amino acid position 147 is highlighted (in white and with an arrow) and is present in A/Brevig Mission/1/1918 but not A/Puerto Rico/8/1934.

Figure 9: Cyclical activity of the disrupted peptide sequence of a site of limited variability. Disrupted peptide sequences taken from yearly consensus sequences can be groups according to their chemical properties. If arranged with time, the cyclical nature of the disrupted peptide sequence becomes apparent. Other ways of arranging the sequences based on their chemical properties are possible and this is simply one possible incarnation and representative of the cyclical nature of this epitope region.

Figure 10: Amino acid changes at position 147 cycle between four possibilities. A. The identity of amino acid at position 147 cycles between lysine, arginine, isoleucine and is absent for five periods between 1918-1957 and 1977-2015.

EXAMPLES

The present invention is further illustrated by the following Examples, in which parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should

be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, various modifications of the invention in addition to those shown and described herein will be apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims.

10 **Example 1: Antigenic thrift model**

The existence of protective epitopes of low variability is consistent with the population dynamics of influenza A under the “antigenic thrift model”. This model is based on a multi-locus representation of the virus with each locus corresponding to an epitope region and presents an alternative to the more widely accepted “antigenic drift” model in having the potential to contain protective epitopes of limited variability as well as those of high variability. Figure 1 shows how these may locate to the known antigenic sites on a monomer of haemagglutinin (HA). The epidemic behaviour of influenza can be readily explained within the antigenic thrift framework by assuming that most influenza strains are in competition with each other because they share epitopes in regions of low variability (Recker *et al.* 2007; Wikramaratna *et al.* 2013). Thus although new strains may be generated constantly through mutation, most of these cannot expand in the host population due to pre-existing immune responses against their less variable epitopes. This leads to cyclical dominance of antigenic types (Figure 2). By contrast with the “antigenic drift” model, antigenic distance between epidemic strains does not necessarily accumulate with time; instead it periodically expands and contracts.

Carter *et al.* (2013) provides evidence for the antigenic thrift model. Ferrets were infected with one of several historical influenza viruses. Serum antibodies were measured at day 14 and 81 by haemagglutinin inhibition (HAI) assays (Figure 5) and showed a periodic cross-reactivity of antibodies to viruses of historic strains, as predicted by the antigenic thrift hypothesis.

The observed periodic cross-reactivity in Carter *et al.* (2013) is predicted, in part, by the current structural bioinformatics analysis. For example, infection of ferrets with the 1957 strain is predicted to induce cross-reactive antibodies to the A/R/8/1934, A/Den/1/1957, A/NC/20/1999 and A/Bris/59/2007 strains. Cross-reactivity is observed between the A/R/8/1934, A/Den/1/1957 and A/NC/20/1999 strains on infection, but not the A/Bris/59/2007 strain.

Example 2: Cyclical cross-reactivity of infant plasma against chronologically-dispersed H1 influenza strains

Standardised enzyme-linked immunosorbant assays (ELISAs) were performed using plasma from children aged 12 to 17 months, collected in 2012. The HA1 domains from influenza strains A/California/4/2009, A/USSR/90/1977, A/Brevig Mission/1/1918, A/Solomon Islands/3/2006, A/New Caledonia/20/1999, A/Puerto Rico/8/34 and A/WSN/33 were bought from Sino Biological. Standards used sera from adults based on their date of birth. Two negative controls were run on each plate consisting of a caesin only control and a non-reactive human plasma or sera control.

The results are shown in Figure 3. Plasma from 81 children aged 12 to 17 months, collected in 2012, cross-reacted with HA1 domains (the head domain and part of the stem domain of HA H1) from influenza strains A/California/4/2009, A/USSR/90/1977 and A/Brevig Mission/1/1918 but not A/Solomon Islands/3/2006, A/New Caledonia/20/1999, A/Puerto Rico/8/34 or A/WSN/33. All ELISA results were completed in triplicate and normalised to non-reactive human plasma. ELISA results were accepted or declined based on the criteria set out in Miura *et al.* (2008).

The fact that this plasma reacted with a panel of historical H1N1 strains in a cyclical manner lead us to infer that epitopes of limited variability are present in the head domain of H1 HA and that they cycle through a limited number of conformations as host population immunity changes.

Example 3: Identification of epitopes of limited variability

To identify epitopes of limited variability, antibody binding sites were mapped to the A/Puerto Rico/8/1934 crystal structure and the variability within those site determined by referring to an alignment of 2,756 H1 sequences (Figure 7). Only parts of A/Puerto Rico/8/1934 crystal structure accessible to antibody binding were considered. This was determined by aligning the crystal structures of A/Puerto Rico/8/1934, A/Brevig Mission/1/1918 and A/California/04/2009 to determine which residues were present on the surface of the protein and the accessibility of those residues. Typically, the accessibility of the positions were the same in all crystal structures but when a position was more accessible in one crystal structure than another crystal structure, the position was allocated as being more accessible to prevent the false identification of sites of limited variability. Parts of the HA contained within the virion were also not considered for analysis.

In silico analysis was used to determine how the accessibility and binding site area contributed to the variability of hypothetical antibody binding sites. An antibody binding site of 800 Å² was used to determining the variability for three accessibility parameters: amino acids with >30%, >10% or >1% accessibility. A dataset of positions with >10% accessibility was used to determine the variability for three binding site sizes: 600 Å², 800 Å² or 1000 Å². Both approaches identified the same regions of limited variability within the head of H1 HA (Figure 7).

Analysis of the sites of limited variability predicted to exist from the *in silico* analysis was performed by mapping the predicted epitopes to the A/Puerto Rico/8/1934, A/Brevig Mission/1/1918 and A/California/04/2009 crystal structures using Swiss-pdb viewer. By mapping the predicted sites to the crystal structures, sites that were likely to be epitopes could be identified. One site close to the receptor binding site (RBS), in a region which is known to be under strong immune selection but thought to be highly variable, centred on a 800 Å² region surrounding this positions 156/158 (Figure 8; Caton *et al.* 1982).

Example 4: Cycling of epitopes

Yearly consensus sequences were generated by dividing the 12,480 curated H1 HA sequences into separate fasta files based on the year that the sequence was collected.

5 The R package 'seqinr' was then used to generate consensus sequences.

Analysis of the predicted binding site surrounding positions 156/158 indicated that there were a number of positions in which charged residues could be found, in addition to positions with non-charged residues, which were either conserved or changed between
10 similar residue types. It is generally accepted that antibodies preferentially bind to charged residues and so the possible epitope permutations were defined based on the cycling of charged amino acids in positions 147, 156, 157, 158 and 159 (Kringelum *et al.* 2013).

15 At position 147, the amino acid alternated between a positively charged amino acid, lysine or arginine, a neutral amino acid, isoleucine, and no amino acid. Consequently the site was divided up based on this pattern into three groups.

Phylogenetic analysis of position 147 also reveals that strains were identified in which
20 no amino acid was present at position 147 five times during the evolution of H1 influenza in humans between 1918-1957 and 1977-2015 (Figure 10). It was also found to cycle between lysine, arginine, isoleucine and no amino acid. Consequently, the importance to having a vaccine containing both arginine and lysine as positively charged amino acids were highlighted. This also indicated that the site is structurally
25 limited and cycling between a small number of conformations.

The 147 positive group was then further divided based on the presence of a positively charged amino acid at position 158 or 159, 158 and either a positive charged amino acid at position 156 or 157.

- 52 -

The space filling capacity of non-charged amino acid was then considered, which allowed an addition group to be produced from 147-positive/158 or 157 positive group based on whether alanine or asparagine was present at position 156 (Figure 9).

5 **Example 5: Loss of neutralisation upon site-directed mutagenesis**

- Sera from children aged 6 to 11 years, taken in late 2006/early 2007, cross-reacted extensively with HA1 domains from historical influenza strains (Figure 4). Cross-reactivity peaks towards to the HA1 domain from A/WSN/33 in addition to the A/New
- 10 Caledonia/20/1999 HA1 domain which is closely related to the A/Solomon Islands/3/2006 HA1 domain. Cross-reactivity was also observed towards A/California/4/2009, A/USSR/90/1977, A/Albany/12/1951, A/Puerto Rico/8/34 and A/Brevig Mission/1/1918.
- 15 Using a microneutralisation assay (Figure 5), up to a 32-fold loss of neutralisation to A/Solomon Islands/3/2006 pseudotyped lentivirus was observed when a lysine was inserted at position 147 (p-value: 0.0005). Up to a 18.75-fold loss of neutralisation of A/WSN/1933 pseudotyped lentivirus was also observed with the insertion of a lysine at position 147 (p-value of 0.0056). While only a 11 serum samples from the UK cohort
- 20 showed cross-reactivity with A/PR/8/1934 pseudotyped lentivirus, the insertion of a lysine at position 147 caused total loss of neutralisation in 8 samples and a reduction in 3 samples. This indicates that the bulk of cross-reactivity between these strains is mediated through an epitope which contains a deletion in position 147.
- 25 This data emphasises the importance of amino acid position 147. It should be noted that in the A/Solomon Islands/3/2006, A/PR/8/1934 and A/WSN/1933 strains no amino acid is contained at position 147. Instead monomers of these viruses consist of 565 instead of 566 amino acids.

Example 6: Synthesis of polypeptides

Invitrogen® GeneArt Strings were used to synthesise the chimeric HA molecules consisting of the epitope of limited variability substituted into the HA1 domain of H5, H6, or H11. Three conformations of the site were initially used.

The chimeric HA1 domain sequences were then cloned into DNA expression constructs and lentiviral glycoprotein expression vectors. The DNA expression were grown up in *E. coli* and purified using a Qiagen Giga Prep Kit. Lentiviruses were produced displaying the chimeric HAs via the protocol outlined in Carnell *et al.*, (2015) before being purified by sucrose cushion centrifugation. The conformations substituted into the H6, H5 and H11 HA1 domains are provided below (amino acid position is denoted in brackets):

Blue:

N (146), K (147), G (148), V (149), A (151), P (154), H (155), A (156), G (157), A (158), K (159), K (163)

AAC (146) AAG (147) GGC (148) GTG (149) GCC (151) CCC (154) CAC (155) GCC (156) GGC (157) GCC (158) AAG (159) AAG (163)

Hazel:

N (146), I (147), G (148), V (149), A (151), S (154), H (155), A (156), G (157), K (158), S (159), K (163)

AAC (146) ATC (147) GGC (148) GTG (149) GCC (151) AGC (154) CAC (155) GCC (156) GGC (157) AAG (158) AGC (159) AAG (163)

Green:

T (146), R (147), G (148), V (149), A (151), S (154), H (155), K (156), G (157), K (158), S (159), R (163)

- 54 -

ACC (146) AGG (147) GGC (148) GTG (149) GCC (151) AGC (154) CAC (155) AAG
(156) GGC (157) AAG (158) AGC (159) AGG (163)

Orange:

5 T (146), K (147), G (148), V (149), A (151), S (154), H (155), N (156), G (157), K (158),
S (159), R (163)

ACC (146) AAG (147) GGC (148) GTG (149) GCC (151) AGC (154) CAC (155) AAC
(156) GGC (157) AAG (158) AGC (159) AGG (163)

10

Red:

T (146), Absent (147),), G (148), V (149), A (151), S (154), H (155), N (156), G (157), K
(158), S (159), R (163)

15 ACC (146) Absent (147) GGC (148) GTG (149) GCC (151) AGC (154) CAC (155) AAC
(156) GGC (157) AAG (158) AGC (159) AGG (163)

These sequences correspond to those cloned into the vaccine constructs and so any
cross-reactivity can be directly attributed to them.

20

Example 7: Mouse challenges

Mouse influenza challenges are performed with influenza strains:

(i) A/California/4/2009 at a concentration of 1×10^5 Pfu and (ii) A/PR/8/1934 at a
25 concentration of 1×10^3 Pfu . Weight changes were monitored on a daily basis. The
optimisation of challenge experiments enables the protection induced by the vaccine in
mice to be quantified in the vaccination studies.

The basic vaccination protocol is shown in Figure 6A.

30

Mice were sequentially vaccinated with the sequences outlined below:

Name	Position											
	146	147	148	149	151	154	155	156	157	158	159	163
Blue	N	K	G	V	A	P	H	A	G	A	K	K
Hazel	N	I	G	V	A	S	H	A	G	K	S	K
Green	T	R	G	V	A	S	H	K	G	K	S	R
Orange	T	K	G	V	A	S	H	N	G	K	S	R
Red	T	Absent	G	V	A	S	H	N	G	K	S	R

Five groups of six mice (named blue, red, hazel, orange and green) were vaccinated via intramuscular injection of 100 µg of DNA each with a different conformation of the epitope substituted into H6 HA at 10 weeks. At 13 weeks of age, the same groups were vaccinated via intramuscular injection of 100 µg of DNA with the same conformation substituted into H5 HA. At 18 weeks of age, the same groups were vaccinated via intramuscular with the same conformation substituted into H11 HA displayed on a lentivirus and mixed with Alum adjuvant (Alhydrogel, Invivogen). Two control groups (purple and grey) were vaccinated in the same manner as the aforementioned mice with the HAs without the epitope conformations substituted into them. Finally, two further control groups (black and white) were mock vaccinated at 18 weeks with PBS and Alum (Alhydrogel, Invivogen). At 11 weeks, 14 weeks, 20 weeks and 21 weeks all groups were bled. At 22 weeks, the blue, orange, hazel and purple groups were challenged with mouse adapted A/California/4/2009 virus and weighed daily. At 22 weeks, the red, green, grey and white groups were challenged with mouse adapted A/PR/8/1934 virus and weighed daily. The results are shown in Figure 6.

Example 8: Vaccination against the H3 influenza subtype

The method of identifying sites of limited variability and subsequent epitopes of limited variability is applied to the H3 subtype of influenza A. As H3 subtype influenza A virus evolves in a similar way to the H1 subtype influenza A virus, this approach to identifying epitopes is equally applicable to H3 subtype influenza A. Consequently epitopes can be identified by mapping the variability of H3 strains to the head of H3 influenza, identifying regions of limited variability, mapping said regions to H3 structures to identify potential

epitopes and then analysing consensus sequence data to identify epitopes behaving in a cyclical manner predicted by antigenic thrift model

Epitope conformations of this type are placed in the HA head domains of H4, H7 H10, H14 and H15 and expressed using VLPs or viral vectors. The vaccine combination is administered as a prime-boost-boost.

Example 9: Cross-reactivity produced by the vaccines

ELISA assays were performed against the HA1 domain of A/PR/8/1934, A/Bel/1942, A/Albany/14/1951 and A/Memphis/3/1987. Relative ELISA units (REU) were calculated based on a known positive sample reaching an OD of 1.0 in each assay.

Vaccine groups (pooled sera samples)	HA1 domains (REU)			
	A/PR/8/ 1934	A/Bel/1942	A/Albany/ 14/1951	A/Memphis /3/1987
Blue	-	-	-	-
Hazel	-	-	-	329
Green	205	-	-	569
Orange	231	224	317	735
Red	770	-	-	307
H5+H6+H11 control	-	-	-	-
Unvaccinated control	-	-	-	-

REFERENCES

- Belongia, E.A. *et al.*, 2009. Effectiveness of Inactivated Influenza Vaccines Varied Substantially with Antigenic Match from the 2004-2005 Season to the 2006-2007 Season Linked references are available on JSTOR for this article : Effectiveness of Inactivated Influenza Vaccines Varied. *The Journal of Infectious Disease*, 199(2), pp.159–167.
- Carnell *et al.*, (2015) Pseudotype-based neutralization assays for influenza: a systematic analysis. *Front Immunol.* 2015 Apr 29;6:161. doi: 10.3389/fimmu.2015.00161. eCollection 2015.
- Carter *et al.*, (2013) Sequential seasonal H1N1 influenza virus infections protect ferrets against novel 2009 H1N1 influenza virus. *J Virol.* 2013 Feb;87(3):1400-10.
- Caton *et al.*, 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell*, 31(2 Pt 1), pp.417–427.
- Gupta S. 2016 Immune Driven Pathogen Evolution, *Encyclopaedia of Immunology* (Ed. Kaye, P.) Elsevier.
- Krammer, F. *et al.*, 2013. Broadly Protective Stalk-Specific Antibodies. , 87(12), pp.6542–6550.
- Li, Y. *et al.*, 2013. Immune history shapes specificity of pandemic H1N1 influenza antibody responses. 210(8), pp.1493–1500.
- Lozano, R. *et al.*, 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010 : a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, pp. 2095–2128.

Manicassamy, B. *et al.*, 2010. Protection of mice against lethal challenge with 2009 H1N1 influenza A virus by 1918-like and classical swine H1N1 based vaccines. PLoS Pathogens, 6(1).

- 5 Matsuzaki, Y. *et al.*, 2014. Epitope Mapping of the Hemagglutinin Molecule of A / (H1N1) pdm09 Influenza Virus by Using Monoclonal Antibody Escape Mutants. Journal of Virology, 88(21), pp.12364–12373.

- 10 Mertz, D., Hyong, T. & Johnstone, J., 2013. Populations at risk for severe or complicated influenza illness : systematic review and meta-analysis. British Medical Journal, 5061(August), pp.1–15.

Miura *et al.* 2008 *Vaccine* 26:193.

- 15 Presanis, A.M. *et al.*, 2011. Changes in severity of 2009 pandemic A / H1N1 influenza in England : a Bayesian evidence synthesis. British Medical Journal, (343), pp.1–14.

- 20 Recker, M. *et al.*, 2007. The generation of influenza outbreaks by a network of host immune responses against a limited set of antigenic types. PNAS 104:7711

Taubenberger, J.K. & Morens, D.M., 2006. 1918 Influenza : the Mother of All Pandemics. Lancet, 12(1), pp.15–22.

- 25 Treanor, J.J. *et al.*, 2012. Effectiveness of Seasonal Influenza Vaccines in the United States During a Season With Circulation of All Three Vaccine Strains., pp.1–9.

WHO 2016. Recommended composition of influenza virus vaccines for use in the 2016-2017 northern hemisphere influenza season.

- 30 Wikramaratna, P.S. *et al.*, 2013. The antigenic evolution of influenza: drift or thrift? Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 368(1614), p.20120200. Available at:

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3678325&tool=pmcentrez&rendertype=abstract>.

SEQUENCES

5

SEQ ID NO: 1

H1 head domain - amino acid

CKLRGVAPLHLGKCNIA GWILGNPECESLSTASSWSYIVETSSSDNGTCYPGDFIDYEE
LREQLSSVSSFERFEIFPKTSSWPNHDSNKGVT AACPHAGAKSFYKNLIWL VKKGNSY
10 PKLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADAYV FVGTSRYSKKFKPEIAIR
PKVRDQEGRMNYYWTLVEPGDKITFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDC

SEQ ID NO: 2

H6 head domain - amino acid

15 CKILNKAPLDLRGCTIEGWILGNPQCDLLLGDQSWSYIVERPTAQNGICYPGTLNEVEE
LKALIGSGERVERFEMFPKSTWAGVDTNSGVTSACPYNSGSSFYRNLLWIKTKSAAY
PVIKGTYNNTGNQPILYFWGVHHPD TNEQNTLYGSGDRYVRMGTESMNF AKSPEIA
ARPAVNGQRGRIDYYWSVLKPGETLNVESNGNLIAPWYAYKFVSTNNKGAVFKSNLPI
ENC

20

SEQ ID NO: 3

H5 head domain - amino acid

DL DGVKPLILRDCSVAGWLLGNPMCDEF LNVP EWSYIVEKANPANDLCYPGNFNDYE
ELKHLLSRINHF EKI QIIPKSSWSDHE ASSGVSSACPYQGRSSFFRN VVWLIKKNNAYPTI
25 KRSYNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPKIA TRSKV
NGQSGRMEFFWTILKPNDAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGN

SEQ ID NO: 4

H11 head domain - amino acid

30 CSIDGKAPISLGDCSFAGWILGNPMCDDLIGKTSWSYIVEKPNPTNGICYPGTLENEEE
LRLKFSGVLEFSKFEAFTSNGWGAVNSGAGVTAACKFGSSNSFFRN MVWLIHQSGTY

PVIRRTFNNTKGRDVL MVWGVHHPATLKEHQDLYKKDSSYVAVGSESYNRRFTPEIST
RPKVNGQAGRMTFYWTIVKPGEAITFESNGAFLAPRYAFELVSLGNGKLFRSDLNIESC

SEQ ID NO: 5**5 H1 head domain - nucleotide**

TGCAAGCTGAGGGGCGTGGCCCCCTGCACCTGGGCAAGTGCAACATCGCCGGCTGGATC
CTGGGCAACCCCGAGTGCGAGAGCCTGAGCACCGCCAGCAGCTGGAGCTACATCGTGGAG
ACCAGCAGCAGCGACAACGGCACCTGCTACCCCGGCGACTTCATCGACTACGAGGAGCTG
AGGGAGCAGCTGAGCAGCGTGAGCAGCTTCGAGAGGTTGAGATCTTCCCCAAGACCAGC
10 AGCTGGCCCAACCACGACAGCAACAAGGGCGTGACCGCCGCCTGCCCCACGCCGGCGCC
AAGAGCTTCTACAAGAACCTGATCTGGCTGGTGAAGAAGGGCAACAGCTACCCCAAGCTG
AGCAAGAGCTACATCAACGACAAGGGCAAGGAGGTGCTGGTGCTGTGGGGCATCCACCAC
CCCAGCACCGAGCGCCGACCAGCAGAGCCTGTACCAGAACGCCGACGCCCTACGTGTTCTG
GGCACCAGCAGGTACAGCAAGAAGTTCAAGCCCGAGATCGCCATCAGGCCCAAGGTGAGG
15 GACCAGGAGGGCAGGATGAACTACTACTGGACCCTGGTGGAGCCCGGCGACAAGATCACC
TTCGAGGCCACCGGCAACCTGGTGGTGCCAGGTACGCCTTCGCCATGGAGAGGAACGCC
GGCAGCGGCATCATCATCAGCGACACCCCCGTGCACGACTGC

SEQ ID NO: 6**20 H6 head domain - nucleotide**

TGCAAGATCCTGAACAAGGCCCCCTGGACCTGAGGGGCTGCACCATCGAGGGCTGGATC
CTGGGCAACCCCGAGTGCGACCTGCTGCTGGGCGACCAGAGCTGGAGCTACATCGTGGAG
AGGCCCACCGCCAGAACGGCATCTGCTACCCCGGCACCCTGAACGAGGTGGAGGAGCTG
AAGGCCCTGATCGGCAGCGGCGAGAGGGTGGAGAGGTTGAGATGTTCCCCAAGAGCACC
25 TGGGCGCGCGTGAGACACCAACAGCGGCGTGACCAGCGCCTGCCCCCTACAACAGCGGCAGC
AGCTTCTACAGGAACCTGCTGTGGATCATCAAGACCAAGAGCGCCGCCTACCCCGTGATC
AAGGGCACCTACAACAACACCGGCAACCAGCCCATCCTGTACTTCTGGGGCGTGACCCAC
CCCCCGACACCAACGAGCAGAACACCCTGTACGGCAGCGGCGACAGGTACGTGAGGATG
GGCACCGAGAGCATGAACTTCGCCAAGAGCCCCGAGATCGCCGCCAGGCCCGCCGTGAAC
30 GGCCAGAGGGGCGAGGATCGACTACTACTGGAGCGTGCTGAAGCCCGGCGAGACCCTGAAC
GTGGAGAGCAACGGCAACCTGATCGCCCCCTGGTACGCCTACAAGTTCGTGAGCACCAAC
AACAAGGGCGCCGTGTTCAAGAGCAACCTGCCCATCGAGAACTGC

SEQ ID NO: 7**35 H5 head domain - nucleotide**

TGCGACCTGGACGGCGTGAAGCCCCTGATCCTGAGGGACTGCAGCGTGGCCGGCTGGCTG
CTGGGCAACCCCATGTGCGACGAGTTCCTGAACGTGCCCGAGTGGAGCTACATCGTGGAG
AAGGCCAACCCCGCCAACGACCTGTGCTACCCCGGCAACTTCAACGACTACGAGGAGCTG
AAGCACCTGCTGAGCAGGATCAACCACTTCGAGAAGATCCAGATCATCCCCAAGAGCAGC
40 TGGAGCGACACGAGGCCAGCAGCGGCGTGAGCAGCGCCTGCCCCCTACCAGGGCAGGAGC
AGCTTCTTCAGGAACGTGGTGTGGCTGATCAAGAAGAACAACGCCTACCCACCATCAAG
AGGAGCTACAACAACACCAACCAGGAGGACCTGCTGGTGCTGTGGGGCATCCACCACCCC
AACGACGCCGCCGAGCAGACCAAGCTGTACCAGAACCCACACCTACATCAGCGTGGGC

- 61 -

ACCAGCACCCCTGAACCAGAGGCTGGTGCCCAAGATCGCCACCAGGAGCAAGGTGAACGGC
 CAGAGCGGCAGGATGGAGTTCTTCTGGACCATCCTGAAGCCCAACGACGCCATCAACTTC
 GAGAGCAACGGCAACTTCATCGCCCCCGAGTACGCCTACAAGATCGTGAAGAAGGGCGAC
 AGCACCATCATGAAGAGCGAGCTGGAGTACGGCAACTGC

5

SEQ ID NO: 8**H11 head domain - nucleotide**

TGCAGCATCGACGGCAAGGCCCCCATCAGCCTGGGCGACTGCAGCTTCGCCGGCTGGATC
 CTGGGCAACCCCATGTGCGACGACCTGATCGGCAAGACCAGCTGGAGCTACATCGTGGAG
 10 AAGCCCAACCCCAACCAACGGCATCTGCTACCCCGGCACCCTGGAGAACGAGGAGGAGCTG
 AGGCTGAAGTTCAGCGGCGTGCTGGAGTTCAGCAAGTTCGAGGCCTTCACCAGCAACGGC
 TGGGGCGCCGTGAACAGCGGCGCCGGCGTGACCGCCGCCCTGCAAGTTCGGCAGCAGCAAC
 AGCTTCTTCAGGAACATGGTGTGGCTGATCCACCAGAGCGGCACCTACCCCGTGATCAGG
 AGGACCTTCAACAACACCAAGGGGCAGGGACGTGCTGATGGTGTGGGGCGTGCAACCACCCC
 15 GCCACCCTGAAGGAGCACCAGGACCTGTACAAGAAGGACAGCAGCTACGTGGCCGTGGGC
 AGCGAGAGCTACAACAGGAGGTTCACCCCGGAGATCAGCACCAGGCCCAAGGTGAACGGC
 CAGGCCGGCAGGATGACCTTCTACTGGACCATCGTGAAGCCCGGCGAGGCCATCACCTTC
 GAGAGCAACGGCGCCTTCCTGGCCCCCAGGTACGCCTTCGAGCTGGTGAGCCTGGGCAAC
 GGCAAGCTGTTCAGGAGCGACCTGAACATCGAGAGCTGC

20

SEQ ID NO: 9**H1 haemagglutinin – amino acid**

MKAILVLLYTFATANADTLCIGYHANNSTDVDTVLEKNVTVTHSVNLLEDKHNGKLC
 KLRGVAPLHLGKCNIAGWILGNPECESLSTASSWSYIVETSSSDNGTCYPGDFIDYEEL
 25 REQLSSVSSFERFEIFPKTSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYP
 KLSKSYINDKGKEVLVLWGIHHPSTSADQQSLYQNADAYVFGTSRYSKKFKPEIAIR
 PKVRDQEGRMNYYWTLVEPGDKITFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDC
 NTTCTPKGAINTSLPFQNIHPITIGKCPKYVKSTKLRLATGLRNVPSIQSRGLFGAIAFG
 IEGGWTGMVDGWYGYHHQNEQGSGYAADLKSTQNAIDEITNKVNSVIEKMNTQFTAV
 30 GKEFNHLEKRIENLNKKVDDGFLDIWTYNAELLVLLNERTLDYHDSNVKNLYEKVRS
 QLKNNAKEIGNGCFEFYHKCDNTCMESVKNGTYDYPKYSEEAKLNREEIDGVKLESTR
 IYQILAIYSTVASSLVLVVSLGAISFWMCSNGSLQCRICI

SEQ ID NO: 10**35 H6 haemagglutinin – amino acid**

MIAIIVAILAATGKSDKICIGYHANNSTTQVDITLEKNVTVTHSVVELLENQKEERFCKILNK
 APLDLRGCTIEGWILGNPQCDLLLGDQSWSYIVERPTAQNGICYPGTLNEVEELKALIG

- 62 -

SGERVERFEMFPKSTWAGVDTNSGVTSA CPYNSGSSFYRNLLWIKTKSAAYPVIKGT
 YNNTGNQPILYFWGVHHPPDTNEQNTLYGSGDRYVRMGTESMNFAKSPEIAARPAVN
 GQRGRIDYYWSVLKPGETLNVESNGNLIAPWYAYKFVSTNNKGAVFKSNLPIENC DAT
 CQTIAGVLR TNKTFQNV SPLWIGEC PKYVKSESLRLATGLRNVPQIETRGLFGA IAGFIE
 5 GGWTGMIDGWYGYHHENSQGS GYAADRESTQKAIDGITNKVNSIIDKMNTQFEAVDH
 EFSNLERRIDNLNKR MEDGFLDVWTYNAELLV LLENERTLDLHDANVKNLYEKVKS QL
 RDNANDLGNGCFEFWHKCDNECIESVKNGTYDYPKYQDESKLNRQEIESV
 KLENLGVYQILAIYSTVSSSLVLVGLIAMGLWMCSNGSMQCRICI

10 **SEQ ID NO: 11****H5 haemagglutinin – amino acid**

MEKIVLLLAI VSLVKSDQICIGYHANNSTEQVD TIMEKNTVTHA QDILEKTHNGKL~~DL~~
 DGVKPLILRDCSVAGWLLGNPMCDEFLNVPEWSYIVEKANPANDLCYPGNFNDYEELK
 HLLSRINHFEKIQIIPKSSWSDHEASSGVSSACPYQGRSSFFRN VVWLIKKNNAYPTIKRS
 15 YNNTNQEDLLVLWGIHHPNDAAEQTKLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQ
 SGRMEFFWTILKPNDAINFESNGNFI APEYAYKIVKKGDSTIMKSELEYGN~~NT~~KCQTPI
 GAINSSMPFHNIHPLTIGEC PKYVKS NRLVLATGLRN SPQRKKRGLFGA IAGFIEGGWQG
 MVDGWYGYHHSNEQGSGYAADKESTQKAIDGVTNKVNSIIDKMNTQFEAVGREFN NL
 ERRIENLNKKMEDGFLDVWTYNAELLVLMENERTLDFHDSNVKNLYDKVRLQLRDNA
 20 KELGNGCFEFYHKCDNECMESVRNGTYDYPQYSEEARLKREEISGVKLESIGTYQILSIY
 STVASSLALAIMVAGLSLWMCSNGSLQCRICI

SEQ ID NO: 12**H11 haemagglutinin – amino acid**

MKKTLLFAAIIICIQADEICIGYLSNNSTEKVD TIIESNVTVTSSVELVENEHTGSFC SIDGK
 APISLGDCSFAGWILGNPMCDDLIGKTSWSYIVEKPNPTNGICYPGTLENEEELRLKFS
 GVLEFSKF EAFTSNGWGAVNSGAGVTAACKFGSSNSFFRN MVWLIHQSGTYPVIRRT
 FNNTKGRDVLMVWGVHHPATLKEHQDLYKKDSSYVAVGSESYNRRFTPEISTRPKVN
 GQAGRMTFYWTIVKPGEAITFESNGAFLAPRYAFELVSLGNGKLFRSDLNIESCSTKCQ
 30 SEIGGIN TNRSFHNVHRNTIGDCPKYVNVKSLK LATGLRNVP AIATRGLFGA IAGFIEGG
 WPGLINGWYGFQHRNEEGTGIAADKESTQKAIDQITSKVNNIVDRMNTNFESVQHEFS
 EIEERINQLSKHVDDSVIDIWSYNAQLLV LLENENEKTLDLHDSNVRNLHEKVRRMLKDNA
 KDEGNGCFTFYHKCDNECIEKVRNGTYDHKEFEEESKLNRQEIEGVKLD SNGNVYKIL
 SIYSCIASSLVLA AIIIMGFILWACSN GSCRCTICI

35

SEQ ID NO: 13**Head domain (blue) - amino acid BLUE SEQUENCE PLACED INTO THE H11 HEAD**

■SIDGKAPISLGDCSFAGWILGNPECEDLIGKTSWSYIVEKPNPTNGICYPGTLENEEEL
RLKFSGVLEFSKFEAFTSNGWGAVNS NRGV TAAC PHAGAK

5 SFFKNMVWLIHQSGTYPVIRRTFNNTKGRDVLMMVWGVHHPATLKEHQDLYKKDSSYV
AVGSESYNRRFTPEISTRPKVNGQAGRMTFYWTIVKPGEAITFESNGAFLAPRYAFELV
SLGNGKLFRSDLNIES■

SEQ ID NO: 14

10 **Head domain (hazel) - amino acid HAZEL SEQUENCE PLACED INTO THE H6
HEAD DOMAIN**

■KILNKAPLDLRGCTIEGWILGNPECELLLGDQSWSYIVERPTAQNGICYPGTLNEVEE
LKALIGSGERVERFEMFPKSTWAGVDT NIGV TAAC SHAGKS

SFYKNLLWIKTKSAAYPVIKGTYNNTGNQPILYFWGVHHPDPTNEQNTLYGSGDRYVR
15 MGTESMNFAKSPEIAARPAVNGQRGRIDYYWSVLKPGETLNVESNGNLIAPWYAYKF
VSTNNKGAVFKSNLPIEN■

SEQ ID NO: 15**Head domain (green) - amino acid GREEN SEQUENCE PLACED INTO H6 HEAD**

20 **DOMAIN.**

CKILNKAPLDLRGCTIEGWILGNPECELLLGDQSWSYIVERPTAQNGICYPGTLNEVEE
LKALIGSGERVERFEMFPKSTWAGVDTTRGVTAACSHKGKSSFYKNLLWIKTKSAAYP
VIKGTYNNTGNQPILYFWGVHHPDPTNEQNTLYGSGDRYVRMGTESMNFAKSPEIAA
RPAVNGQRGRIDYYWSVLKPGETLNVESNGNLIAPWYAYKFVSTNNKGAVFKSNLPIE

25 NC

SEQ ID NO: 16**Head domain (orange) - amino acid**

■SIDGKAPISLGDCSFAGWILGNPECEDLIGKTSWSYIVEKPNPTNGICYPGTLENEEEL
30 RLKFSGVLEFSKFEAFTSNGWGAVNS TKGV TAAC SHNGKS

SFFRNMVWLIHQSGTYPVIRRTFNNTKGRDVLMMVWGVHHPATLKEHQDLYKKDSSYV

AVGSESYNRRFTPEISTRPKVNGQAGRMTFYWTIVKPGEAITFESNGAFLAPRYAFELV
SLGNGKLFRSDLNIES

SEQ ID NO: 17

5 Head domain (red) - amino acid

DLDGVKPLILRDCSVAGWLLGNPECEEFLNVPEWSYIVEKANPANDLCYPGNF
NDYEELKHLLSRINHFEEKIQIIPKSSWSDHET-
GGVSAACASHNGKSSFFRNVVWLIKKNNAYPTIKRSYNNTNQEDLLVLWGIHHP
NDAAEQTKLYQNPTTYISVGTSTLNQRLVPKIATRSKVNGQSGRMEFFWTILKPN
DAINFESNGNFIAPEYAYKIVKKGDSTIMKSELEYGN

SEQ ID NO: 18

Head domain (blue) – nucleotide

TGCAGCATCGACGGCAAGGCCCCCATCAGCCTGGGCGACTGCAGCTTCGCCGGCTGGATC
CTGGGCAACCCCGAGTGCGAGGACCTGATCGGCAAGACCAGCTGGAGCTACATCGTGGAG
AAGCCCAACCCCAACCAACGGCATCTGCTACCCCGGCACCCTGGAGAACGAGGAGGAGCTG
AGGCTGAAGTTCAGCGGCGTGCTGGAGTTCAGCAAGTTCGAGGGCCTTCACCAGCAACGGC
TGGGGCGCCGTGAACAGCAACAGGGGCGTGACCGCCGCCTGCCCCACGCCGGCGCCAAG
AGCTTCTTCAAGAACATGGTGTGGCTGATCCACCAGAGCGGCACCTACCCCGTGATCAGG
AGGACCTTCAACAACACCAAGGGCAGGGACGTGCTGATGGTGTGGGGCGTGACCCACCCC
GCCACCCTGAAGGAGCACCAGGACCTGTACAAGAAGGACAGCAGCTACGTGGCCGTGGGC
AGCGAGAGCTACAACAGGAGGTTCACCCCGAGATCAGCACCAGGCCCAAGGTGAACGGC
CAGGCCGGCAGGATGACCTTCTACTGGACCATCGTGAAGCCCGGCGAGGCCATCACCTTC
GAGAGCAACGGCGCCTTCCTGGCCCCCAGGTACGCCTTCGAGCTGGTGAGCCTGGGCAAC
GGCAAGCTGTTCAGGAGCGACCTGAACATCGAGAGCTGC

SEQ ID NO: 19

Head domain (hazel) – nucleotide

TGCAAGATCCTGAACAAGGCCCCCCTGGACCTGAGGGGCTGCACCATCGAGGGCTGGATC
CTGGGCAACCCCGAGTGCGAGCTGCTGCTGGGCGACCAGAGCTGGAGCTACATCGTGGAG
AGGCCCACCGCCCAGAACGGCATCTGCTACCCCGGCACCCTGAACGAGGTGGAGGAGCTG
AAGGCCCTGATCGGCAGCGGCGAGAGGGTGGAGAGGTTCGAGATGTTCCCAAGAGCACC
TGGGCGCGCGTGAGACACCAACATCGGCGTGACCGCCGCCTGCAGCCACGCCGGCAAGAGC
AGCTTCTACAAGAACCTGCTGTGGATCATCAAGACCAAGAGCGCCGCCTACCCCGTGATC
AAGGGCACCTACAACAACACCGGCAACCAGCCCATCCTGTACTTCTGGGGCGTGACCCAC
CCCCCGACACCAACGAGCAGAACACCCTGTACGGCAGCGGCGACAGGTACGTGAGGATG
GGCACCAGAGAGCATGAACTTCGCCAAGAGCCCCGAGATCGCCGCCAGGCCCGCCGTGAAC
GGCCAGAGGGGCGAGGATCGACTACTACTGGAGCGTGCTGAAGCCCGGCGAGACCCTGAAC
GTGGAGAGCAACGGCAACCTGATCGCCCCCTGGTACGCCTACAAGTTCGTGAGCACCAAC
AACAAGGGCGCCGTGTTCAAGAGCAACCTGCCCATCGAGAACTGC

SEQ ID NO: 20**Head domain (green) - nucleotide**

5 TGCAAGATCCTGAACAAGGCCCCCTGGACCTGAGGGGCTGCACCATCGAGGGCTGGATC
CTGGGCAACCCCGAGTGCGAGCTGCTGCTGGGCGACCAGAGCTGGAGCTACATCGTGGAG
AGGCCACCGCCAGAACGGCATCTGCTACCCCGGCACCCTGAACGAGGTGGAGGAGCTG
AAGGCCCTGATCGGCAGCGGCGAGAGGGTGGAGAGGTTCGAGATGTTCCCAAGAGCACC
TGGGCGCGCTGGACACCACAGGGGCGTGACCGCCGCTGCAGCCACAAGGGCAAGAGC
AAGAGCTTCTACAAGAACCTGCTGTGGATCATCAAGACCAAGAGCGCCGCTACCCCGTG
10 ATCAAGGGCACCTACAACAACACCGGCAACCAGCCCATCCTGTACTTCTGGGGCGTGAC
CACCCCCCGACACCAACGAGCAGAACACCCTGTACGGCAGCGGCGACAGGTACGTGAGG
ATGGGCACCGAGAGCATGAACTTCGCCAAGAGCCCCGAGATCGCCGCCAGGCCCGCCGTG
AACGGCCAGAGGGGCGAGGATCGACTACTACTGGAGCGTGCTGAAGCCCGGCGAGACCCTG
AACGTGGAGAGCAACGGCAACCTGATCGCCCCCTGGTACGCCTACAAGTTCGTGAGCACC
15 AACACAAGGGCGCCGTGTTCAAGAGCAACCTGCCCATCGAGAACTGC

SEQ ID NO: 21**Head domain (orange) – nucleotide**

20 TGCAGCATCGACGGCAAGGCCCCCATCAGCCTGGGCGACTGCAGCTTCGCCGGCTGGATC
CTGGGCAACCCCGAGTGCGAGGACCTGATCGGCAAGACCAGCTGGAGCTACATCGTGGAG
AAGCCCAACCCACCAACGGCATCTGCTACCCCGGCACCCTGGAGAACGAGGAGGAGCTG
AGGCTGAAGTTCAGCGGCGTGCTGGAGTTCAGCAAGTTCGAGGCCTTCACCAGCAACGGC
TGGGGCGCCGTGAACAGCACCAAGGGCGTGACCGCCGCTGCAGCCACAACGGCAAGAGC
AGCTTCTTCAGGAACATGGTGTGGCTGATCCACCAGAGCGGCACCTACCCCGTGATCAGG
25 AGGACCTTCAACAACACCAAGGGCAGGGACGTGCTGATGGTGTGGGGCGTGCAACACCCC
GCCACCCTGAAGGAGCACCAGGACCTGTACAAGAAGGACAGCAGCTACGTGGCCGTGGGC
AGCGAGAGCTACAACAGGAGGTTACCCCCGAGATCAGCACCAGGCCCAAGGTGAACGGC
CAGGCCGGCAGGATGACCTTCTACTGGACCATCGTGAAGCCCGGCGAGGCCATCACCTTC
GAGAGCAACGGCGCCTTCCTGGCCCCCAGGTACGCCTTCGAGCTGGTGAGCCTGGGCAAC
30 GGCAAGCTGTTTCAGGAGCGACCTGAACATCGAGAGCTGC

SEQ ID NO: 22**Head domain (red) - nucleotide**

35 TGCGACCTGGACGGCGTGAAGCCCCTGATCCTGAGGGACTGCAGCGTGGCCGGCTGGCTG
CTGGGCAACCCCGAGTGCGAGGAGTTCCTGAACGTGCCCCGAGTGGAGCTACATCGTGGAG
AAGGCCAACCCCGCCAACGACCTGTGCTACCCCGGCAACTTCAACGACTACGAGGAGCTG
AAGCACCTGCTGAGCAGGATCAACCACTTCGAGAAGATCCAGATCATCCCCAAGAGCAGC
TGGAGCGACCACGAGACCGGCGGCGTGAGCGCCGCTGCGCCAGCCACAACGGCAAGAGC
AGCTTCTTCAGGAACGTGGTGTGGCTGATCAAGAAGAACAACGCCTACCCACCATCAAG
40 AGGAGCTACAACAACACCAACCAGGAGGACCTGCTGGTGTGTGGGGCATCCACCACCCC
AACGACGCCGCCGAGCAGACCAAGCTGTACCAGAACCCACACCTACATCAGCGTGGGC
ACCAGCACCTGAACCAGAGGCTGGTGCCCAAGATCGCCACCAGGAGCAAGGTGAACGGC
CAGAGCGGCAGGATGGAGTTCTTCTGGACCATCCTGAAGCCCAACGACGCCATCAACTTC

- 66 -

GAGAGCAACGGCAACTTCATCGCCCCCGAGTACGCCTACAAGATCGTGAAGAAGGGCGAC
AGCACCATCATGAAGAGCGAGCTGGAGTACGGCAACTGC

SEQ ID NO: 23**5 H9 haemagglutinin – amino acid**

METVSLITILLVTVSNADKICIGYQSTNSTETVDTLTENNVPVTHAKELLHTEHNGML
CATSLGHPLILDCTIEGLIYGNPSCDLLLGGREWSYIVERPSAVNGLCYPGNVENLEEL
RSLFSSARSYQRIQIFPDTIWNVSYSGTSKACSDSFYRSMRWLTQKNNAYPIQDAQYT
NNQGKNILFMWGINHPPTDTAQTNLYTRTDTTTSVATEEINRTFKPLIGPRPLVNGLQG
10 RIDYYWSVLKPGQTLRIRSNGNLIAPWYGHILSGESHGRILKTDLKRGSCTVQCQTEKG
GLNTTLPFQNVSKYAFGNCSKYIGIKSLKLAVGLRNVPSRSSRGLFGAIAGFIEGGWSG
LVAGWYGFQHSNDQGVGMAARDSTQKAIDKITSKVNNIVDKMNKQYEIIDHEFSEVE
TRLNMINNKIDDQIQDIWAYNAELLVLENQKTLDEHDANVNNLYNKVKRALGSNAVED
GKGCFELYHKCDDQCMETIRNGTYNRRKYQEEKLERQKIEGVKLESEGTYKILTIYST
15 VASSLVIAMGFAAFLFWAMSNGSCRCNICI

AMENDED CLAIMS

received by the International Bureau on 19 February 2018 19.02.(2018)

1. An immunogenic composition comprising at least five different virus-like
 5 particles (VLPs), wherein each VLP independently comprises one or more
 homotrimers comprised of polypeptides of SEQ ID NOs: 13-17, optionally together
 with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or
 diluents, as a combined preparation in a form suitable for simultaneous, separate or
 sequential use for treating or preventing influenza A infection.
 - 10 2. An immunogenic composition comprising two or more polypeptides, wherein
 each polypeptide independently comprises a first region of contiguous amino acids,
 wherein:
 - (a) the amino acid sequence of the first region has at least 80% sequence
 15 identity to an influenza A haemagglutinin head domain; and
 - (b) the first region has one or more amino acid substitutions at positions
 which correspond to the following positions in SEQ ID NO: 9:
 - position 83 is E
 - position 85 is a negatively charged amino acid
 - 20 position 146 is T, N, I or A
 - position 147 is a positively charged amino acid, I or is absent
 - position 148 is G
 - position 149 is V
 - position 151 is A
 - 25 position 154 is S or P
 - position 155 is H
 - position 156 is a positively charged amino acid or A or G or N or E
 - position 157 is a positively charged amino acid or A or G
 - position 158 is a positively charged amino acid or A or S or N or C or E
 - 30 position 159 is K or A or S or N or C
 - position 163 is a positively charged amino acid,
- wherein the amino acid sequences of the two or more polypeptides are different,
 and wherein the composition is capable of inducing antibodies in a subject against

an influenza A virus, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents.

3. A composition comprising a polypeptide, wherein the amino acid sequence of
5 the polypeptide comprises a first region, wherein:

(a) the amino acid sequence of the first region has at least 80% sequence
identity to an influenza A subtype H1, H2, H3, H4, H5, H6, H7, H8, H9, H10,
H11, H12, H13, H14, H15, H16, H17 or H18 haemagglutinin head domain,
10 preferably an influenza A subtype H1, H5, H6, H9 or H11 haemagglutinin
head domain; and

(b) the first region has one or more amino acid substitutions at positions
which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
15 position 85 is a negatively charged amino acid
position 146 is T, N, I or A
position 147 is a positively charged amino acid, I or is absent
position 148 is G
position 149 is V
20 position 151 is A
position 154 is S or P
position 155 is H
position 156 is a positively charged amino acid or A or G or N or E
position 157 is a positively charged amino acid or A or G
25 position 158 is a positively charged amino acid or A or S or N or C or E
position 159 is K or A or S or N or C
position 163 is a positively charged amino acid.

4. A composition as claimed in claim 2 or claim 3, wherein the amino acid
30 sequence of the influenza A haemagglutinin head is selected from the group
consisting of SEQ ID NOs: 1-4.

5. A composition as claimed in any one of claims 2 to 4, the amino acid sequence of the first region has at least 85%, 90% or 95% sequence identity to the influenza A haemagglutinin head domain.

5 6. A composition as claimed in any one of claims 2 to 5, wherein:
(b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
position 85 is E or D
10 position 146 is T or N
position 147 is R or K or I or is absent
position 148 is G
position 149 is V
position 151 is A
15 position 154 is S or P
position 155 is H
position 156 is A or G or K or N or E
position 157 is A or G
position 158 is A or K or E
20 position 159 is A, K, C, N or S
position 163 is K or R.

7. A composition as claimed in any one of claims 2 to 5, wherein:
(b) the first region has one or more amino acid substitutions at positions
25 which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
position 85 is a negatively charged amino acid
position 146 is T, N, I or A
position 147 is a positively charged amino acid,
30 position 148 is G
position 149 is V
position 151 is A
position 154 is S or P

position 155 is H
 position 156 is A
 position 157 is G
 position 158 is K or A or S or N or C
 5 position 159 is K or A or S or N or C
 position 163 is a positively charged amino acid.

8. A composition as claimed in claim 7, wherein:
 (b) the first region has one or more amino acid substitutions at positions
 10 which correspond to the following positions in SEQ ID NO: 9:
 position 83 is E
 position 85 is E
 position 146 is N
 position 147 is R
 15 position 148 is G
 position 149 is V
 position 151 is A
 position 154 is P
 position 155 is H
 20 position 156 is A
 position 157 is G
 position 158 is A
 position 159 is K
 position 163 is K,

25

preferably, all of the above amino acid substitutions.

9. A composition as claimed in any one of claims 2 to 5, wherein:
 (b) the first region has one or more amino acid substitutions at positions
 30 which correspond to the following positions in SEQ ID NO: 9:
 position 83 is E
 position 85 is a negatively charged amino acid
 position 146 is N

position 147 is I
 position 148 is G
 position 149 is V
 position 151 is A
 5 position 154 is S
 position 155 is H
 position 156 is K or A
 position 157 is G
 position 158 is A or K
 10 position 159 is K or S
 position 163 is a positively charged amino acid.

10. A composition as claimed in claim 9, wherein:
 (b) the first region has one or more amino acid substitutions at positions
 15 which correspond to the following positions in SEQ ID NO: 9:
 position 83 is E
 position 85 is E
 position 146 is N
 position 147 is I
 20 position 148 is G
 position 149 is V
 position 151 is A
 position 154 is S
 position 155 is H
 25 position 156 is A
 position 157 is G
 position 158 is K
 position 159 is S
 position 163 is K,

30

preferably, all of the above amino acid substitutions.

11. A composition as claimed in any one of claims 2 to 5, wherein:

(b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:

- position 83 is E
- position 85 is a negatively charged amino acid
- 5 position 146 is T
- position 147 is a positively charged amino acid
- position 148 is G
- position 149 is V
- position 151 is A
- 10 position 154 is S
- position 155 is H
- position 156 is a positively charged amino acid or A or G
- position 157 is a positively charged amino acid or A or G
- position 158 is K
- 15 position 159 is S or C
- position 163 is a positively charged amino acid.

12. A composition as claimed in claim 11, wherein:

- (b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:
- position 83 is E
 - position 85 is E
 - position 146 is T
 - position 147 is R
 - 25 position 148 is G
 - position 149 is V
 - position 151 is A
 - position 154 is S
 - position 155 is H
 - 30 position 156 is K
 - position 157 is G
 - position 158 is K
 - position 159 is S

position 163 is K,

preferably, all of the above amino acid substitutions.

5 13. A composition as claimed any one of claims 2 to 5, wherein:

(b) the first region has one or more amino acid substitutions at positions
which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is a negatively charged amino acid

10 position 146 is T

position 147 is a positively charged amino acid

position 148 is G

position 149 is V

position 151 is A

15 position 154 is S

position 155 is H

position 156 is N

position 157 is G

position 158 is a positively charged amino acid

20 position 159 is S

position 163 is a positively charged amino acid.

14. A composition as claimed in claim 12, wherein:

(b) the first region has one or more amino acid substitutions at positions
25 which correspond to the following positions in SEQ ID NO: 9:

position 83 is E

position 85 is E

position 146 is T

position 147 is K

30 position 148 is G

position 149 is V

position 151 is A

position 154 is S

position 155 is H
 position 156 is N
 position 157 is G
 position 158 is K
 5 position 159 is S
 position 163 is R,

preferably, all of the above amino acid substitutions.

10 15. A composition as claimed in any one of claims 2 to 5, wherein:
 (b) the first region has one or more amino acid substitutions at positions
 which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
 position 85 is a negatively charged amino acid
 15 position 146 is T or N
 position 147 is absent
 position 148 is G
 position 149 is V
 position 151 is A
 20 position 154 is S or P
 position 155 is H
 position 156 is N
 position 157 is G
 position 158 is K
 25 position 159 is S
 position 163 is a positively charged amino acid.

16. A composition as claimed in claim 15, wherein:
 (b) the first region has one or more amino acid substitutions at positions
 30 which correspond to the following positions in SEQ ID NO: 9:

position 83 is E
 position 85 is E
 position 146 is T

position 147 is absent
 position 148 is G
 position 149 is V
 position 151 is A
 5 position 154 is S
 position 155 is H
 position 156 is N
 position 157 is G
 position 158 is K
 10 position 159 is S
 position 163 is R,

preferably, all of the above amino acid substitutions.

15 17. A composition as claimed in claim 2 or claim 3, wherein the first regions of the two or more polypeptides independently comprise amino acid sequences selected from the group consisting of SEQ ID NOs: 13-17.

18. A composition as claimed in any one of the preceding claims, wherein one or
 20 more of the polypeptides additionally comprises a stretch of contiguous amino acids which are derived from a haemagglutinin N-terminal stalk region.

19. A composition as claimed in any one of the preceding claims, wherein the polypeptides are all 280-300 amino acids in length.

25 20. A composition as claimed in any one of the preceding claims, wherein the composition comprises 2, 3, 4 or 5 of said polypeptides, all of which are different.

21. A composition as claimed in any one of the preceding claims, wherein the
 30 composition comprises one or more hetero-trimers of three different polypeptides or homotrimers of the same polypeptides.

22. A kit comprising one or more nucleic acid molecules which code for two or more of the polypeptides as defined in any one of claims 1 to 21.
23. A kit comprising one or more vectors, the one or more vectors comprising
5 nucleic acid molecules which code for one, two or more of the polypeptides as defined in any one of claims 1 to 21.
24. A kit as claimed in claim 23, wherein the vectors are viral vectors, preferably adenoviral vectors or a Modified Vaccinia Ankara (MVA) viral vectors.
- 10 25. A composition comprising a kit as claimed in any one of claims 22 to 24, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents.
- 15 26. A virus-like particle comprising two or more polypeptides as defined in any one of claims 1 to 21.
27. A composition comprising a virus-like particle of claim 26, optionally together with one or more pharmaceutically-acceptable carriers, adjuvants, excipients or
20 diluents.
28. A vaccine composition comprising a composition as claimed in any one of claims 1 to 21, 25 or 27, and an adjuvant.
- 25 29. A composition comprising two or more polypeptides as claimed in any one of claims 1 to 21, 25 or 27 or a vaccine composition as claimed in claim 28 as a combined preparation in a form suitable for simultaneous, separate or sequential use, preferably for treating or preventing influenza A infection.
- 30 30. A composition as claimed in any one of claims 1 to 21, 25 or 27 or a vaccine composition as claimed in claim 28 for use in therapy or for use as a medicament.
31. A composition as claimed in any one of claims 1 to 21, 25 or 27 or a vaccine composition as claimed in claim 28, for use:

- (i) in a method of preventing or treating influenza infection in a subject; or
- (ii) in a method of inducing a T-cell or B-cell response to an influenza antigen in a subject.

5 32. Use of a composition as claimed in any one of claims 1 to 21, 25 or 27 or a vaccine composition as claimed in claim 28, in the manufacture of a medicament for:

- (i) preventing or treating influenza infection in a subject; or
- (ii) inducing a T-cell or B-cell response to an influenza antigen in a subject.

10

33. A method of:

- (i) preventing or treating influenza infection in a subject; or
- (ii) inducing a T-cell or B-cell response to an influenza antigen in a subject,

the method comprising administering an effective amount of a composition as
 15 claimed in any one of claims 1 to 21, 25 or 27 or a vaccine composition as claimed in claim 28 to a subject in need thereof.

34. A method of preventing or treating an influenza infection in a subject or of inducing a T-cell or B-cell response to an influenza antigen in a subject,
 20 the method comprising the steps of:
 (i) simultaneously, separately or sequentially administering an effective amount of one, two, three, four, five or more different polypeptides to a subject in need thereof, wherein each polypeptide independently comprises a first region of contiguous amino acids, wherein:

- 25 (a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and
- (b) the first region has one or more amino acid substitutions at positions which correspond to the following positions in SEQ ID NO: 9:
 position 83 is E
 30 position 85 is a negatively charged amino acid
 position 146 is T, N, I or A
 position 147 is a positively charged amino acid, I or is absent
 position 148 is G
 position 149 is V

position 151 is A

position 154 is S or P

position 155 is H

position 156 is a positively charged amino acid or A or G or N or E

5 position 157 is a positively charged amino acid or A or G

position 158 is a positively charged amino acid or A or S or N or C or E

position 159 is K or A or S or N or C

position 163 is a positively charged amino acid,

10 preferably wherein the first region is defined as in any one of claims 4 to 17.

35. A method as claimed in claim 34, wherein one or more of the polypeptides are in the form of one or more trimers.

15 36. A method as claimed in claim 34 or 35, wherein the method comprises the additional steps of:

(ii) administering a boost with a second polypeptide or second trimer to the subject; and optionally also

20 (iii) administering a boost with a third polypeptide or third trimer to the subject,

wherein the second and third polypeptides/trimers are preferably different to each other and preferably different to the first polypeptide/trimer.

37. A method as claimed in any one of claims 34 to 36, wherein the first, second
25 and third polypeptides are independently selected from the group consisting of polypeptides comprising or consisting of SEQ ID NOs: 13-17 or the first, second and third trimers independently comprise polypeptides comprising or consisting of SEQ ID NOs: 13-17.

30 38. A method as claimed in any one of claims 34 to 37, wherein polypeptides or homotrimers of SEQ ID NOs: 14 and 15 are first administered to the subject; polypeptides or homotrimers of SEQ ID NOs: 13 and 16 are next administered to the subject; and polypeptides or homotrimers of SEQ ID NO: 17 are then administered to the subject.

39. A method as claimed in any one of claims 34 to 38, wherein the polypeptides or trimers are administered to the subject in the form of a VLP.

5 40. A method as claimed in any one of claims 34 to 38, wherein a nucleic acid molecule encoding one or more of the polypeptides is administered to the subject instead of the corresponding polypeptide, preferably wherein the nucleic acid molecule is a viral vector.

10 41. A method for identifying an epitope on a haemagglutinin head domain of an influenza virus of a defined subtype, the method comprising the steps of:

(i) identifying a possible antibody binding site on a haemagglutinin head domain polypeptide of an influenza virus of a defined subtype;

15

(ii) identifying one or more continuous or discontinuous stretches of the head domain polypeptide within the possible antibody binding site; and

20

(iii) comparing the amino acid sequences of those stretches with amino acid sequences of haemagglutinin head domains from a plurality of influenza strains of the defined subtype in order to identify regions of limited amino acid sequence variability within those stretches;

25 thereby identifying a set of positions which have limited variability in their amino acid composition within the haemagglutinin head domain of the influenza virus of the defined subtype and which form an epitope.

42. A process for producing a polypeptide, the process comprising the steps of:

30

(i) using a method for identifying an epitope as claimed in claim 41 to identify a set of amino acids with limited variability within a haemagglutinin head domain of an influenza virus of a defined subtype;

(ii) producing a polypeptide which comprises a first region of contiguous amino acids, wherein:

- 5 (a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and
- (b) the first region has one or more amino acid substitutions at positions which correspond to the positions of the amino acid positions with limited variability, and wherein the substitution introduces the amino acid which is conserved at that position;

10

wherein the polypeptide is capable of inducing antibodies in a subject against an influenza A virus.

43. A process for producing an immunogenic composition, the process
- 15 comprising the steps of:

(i) using a method for identifying an epitope as claimed in claim 41 to identify a set of amino acid positions with limited variability within a haemagglutinin head domain of an influenza virus of a defined subtype;

20

(ii) producing a polypeptide which comprises a first region of contiguous amino acids, wherein:

- (a) the amino acid sequence of the first region has at least 80% sequence identity to an influenza A haemagglutinin head domain; and
- 25 (b) the first region has one or more amino acid substitutions at positions which correspond to the positions of the conserved amino acids, and wherein the substitution introduces the amino acid which is conserved at that position; wherein the polypeptide is capable of inducing antibodies in a subject against an influenza A virus; and

30

(iii) formulating one or more of the polypeptides into an immunogenic composition, wherein the composition optionally comprises one or more pharmaceutically-acceptable carriers, adjuvants, excipients or diluents.

44. A method as claimed in claim 35, wherein one or more of the polypeptides are in the form of one or more homotrimers.

5 ...

- 1 / 29 -

Figure 1

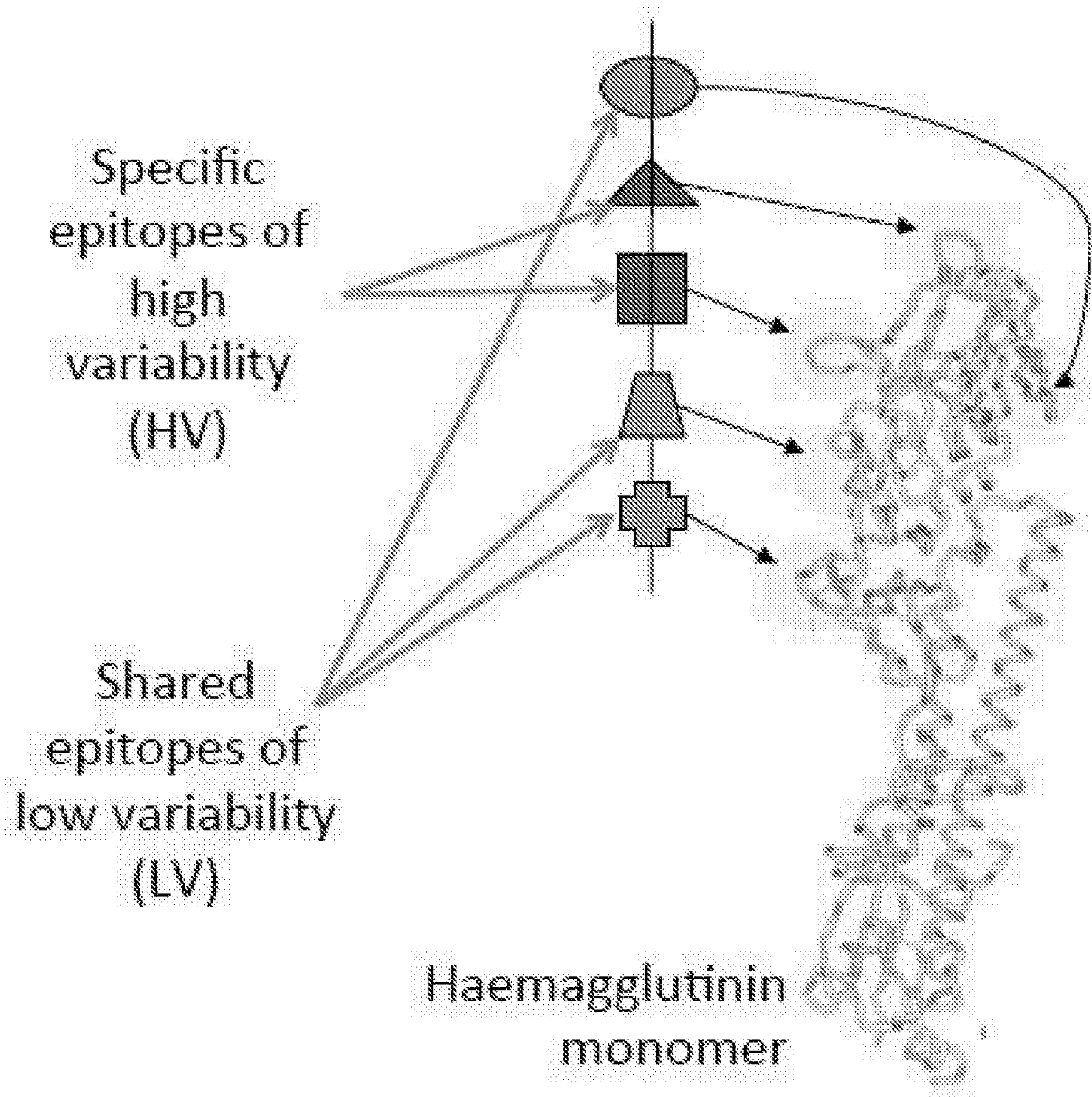


Figure 2

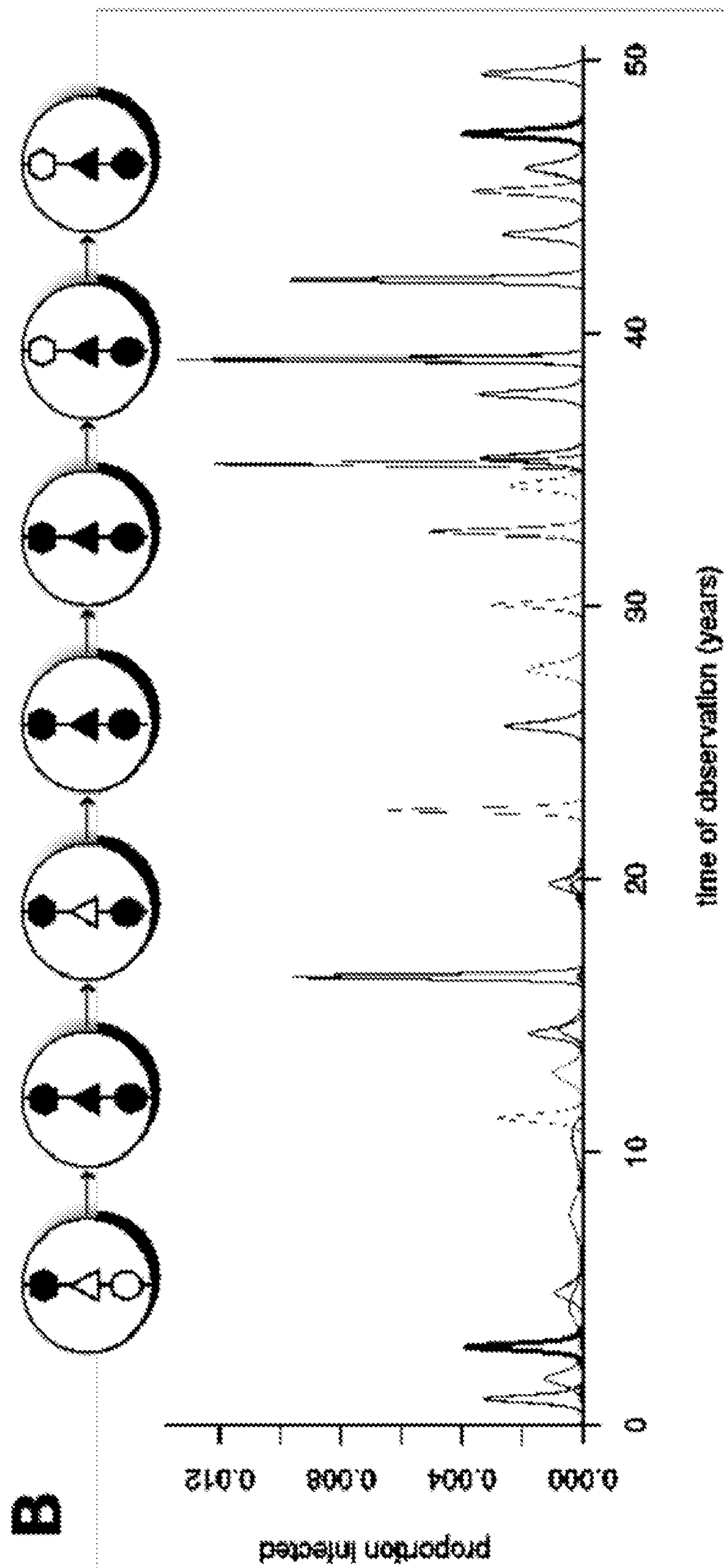


Figure 3

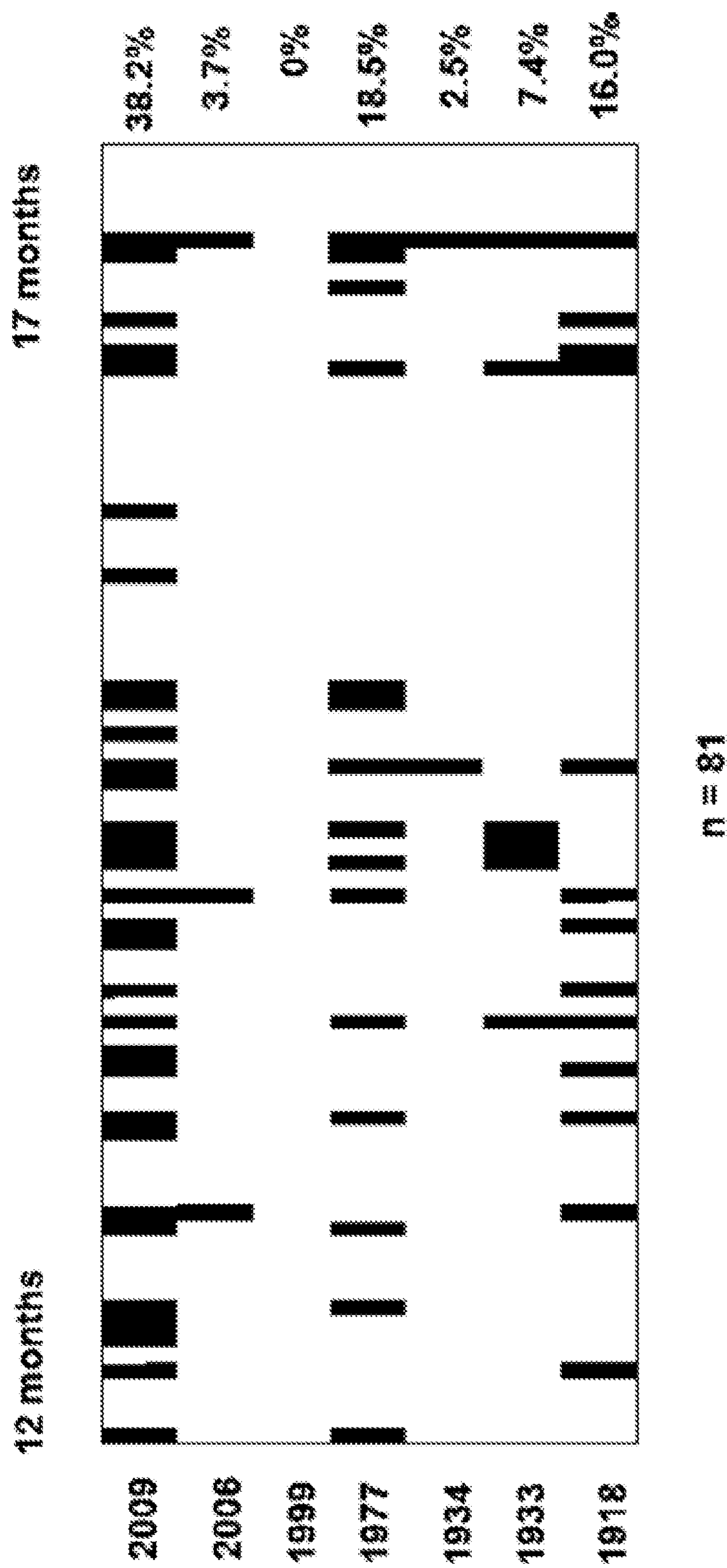


Figure 4



- 5 / 29 -

Figures 5A and 5B

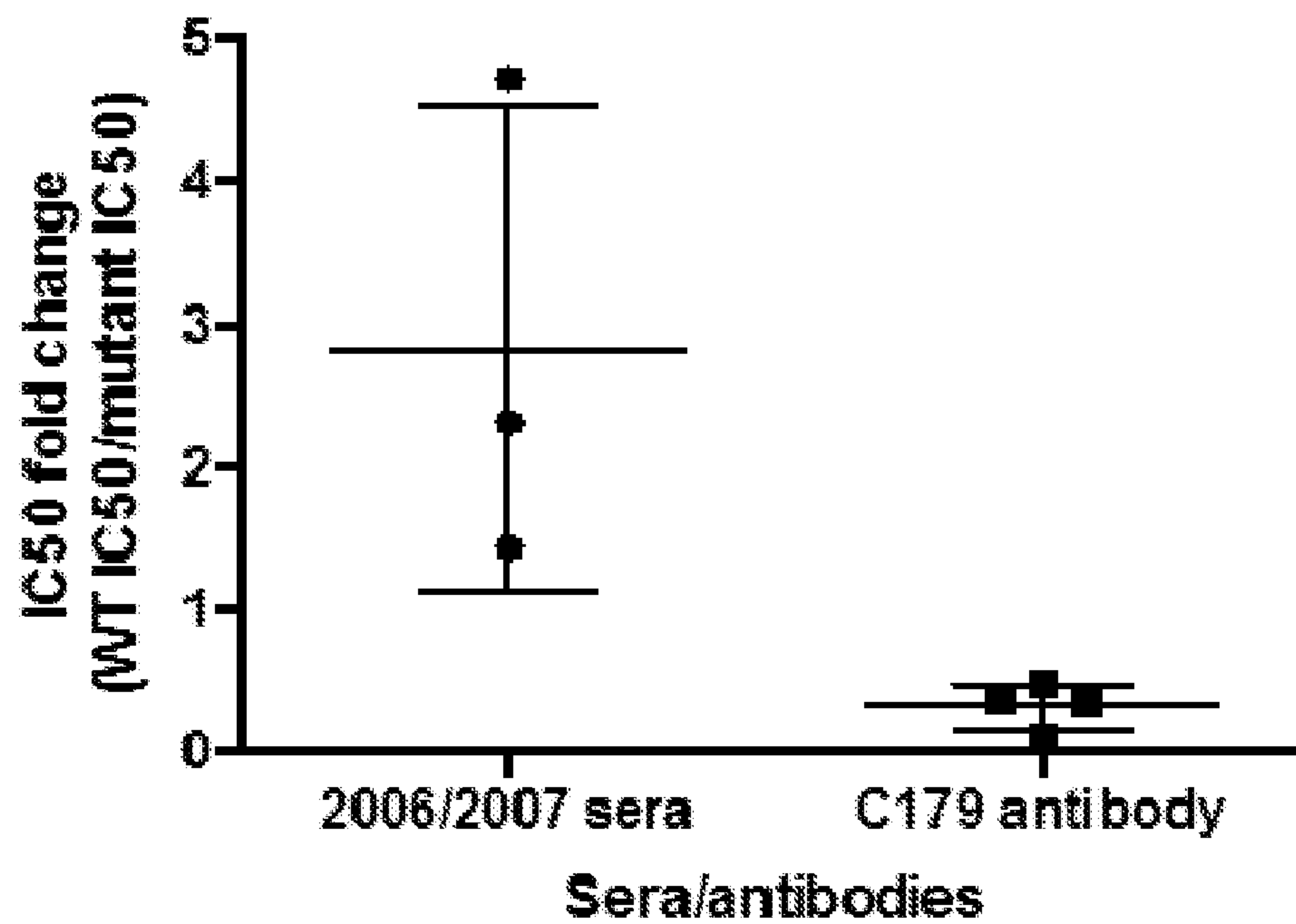
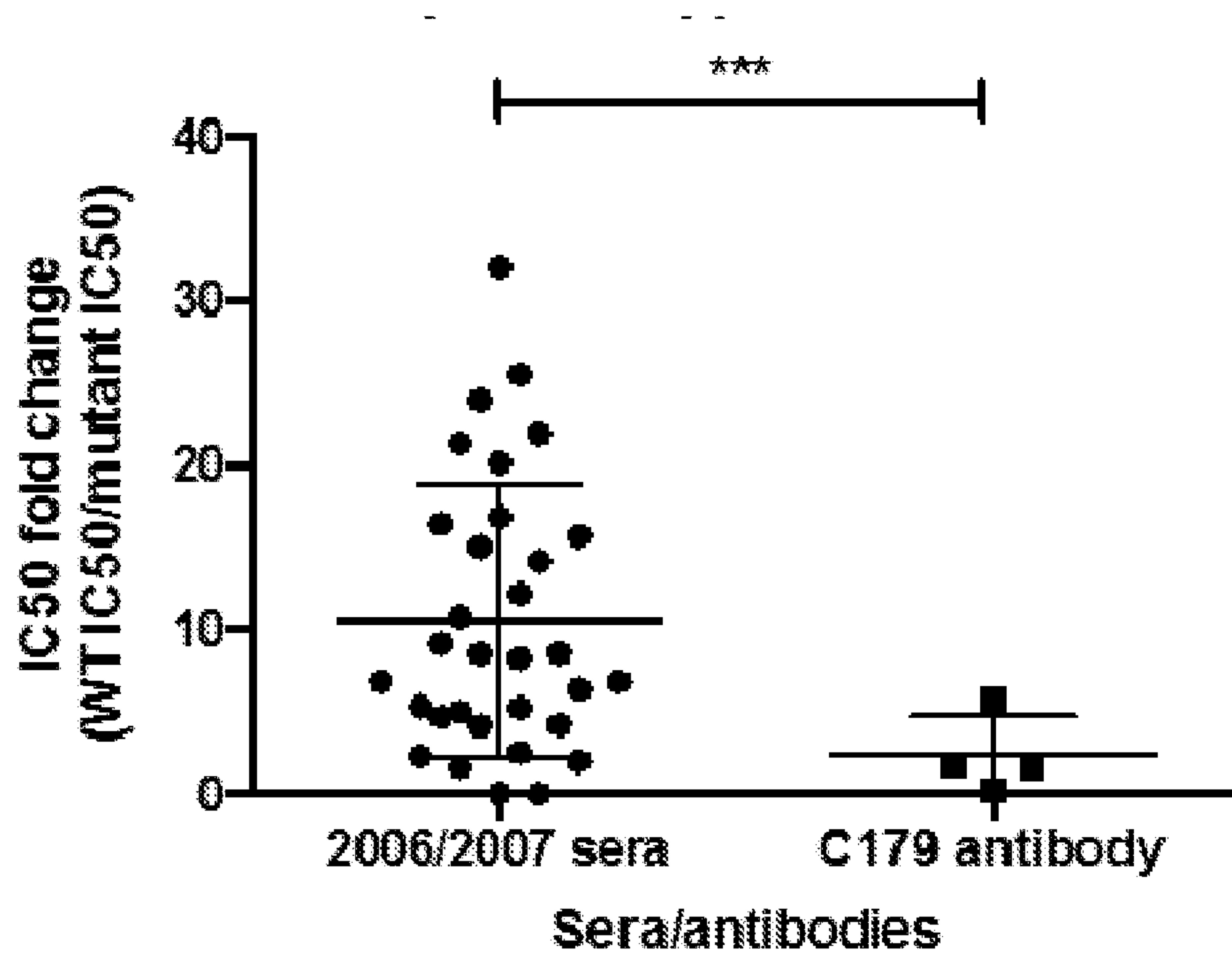
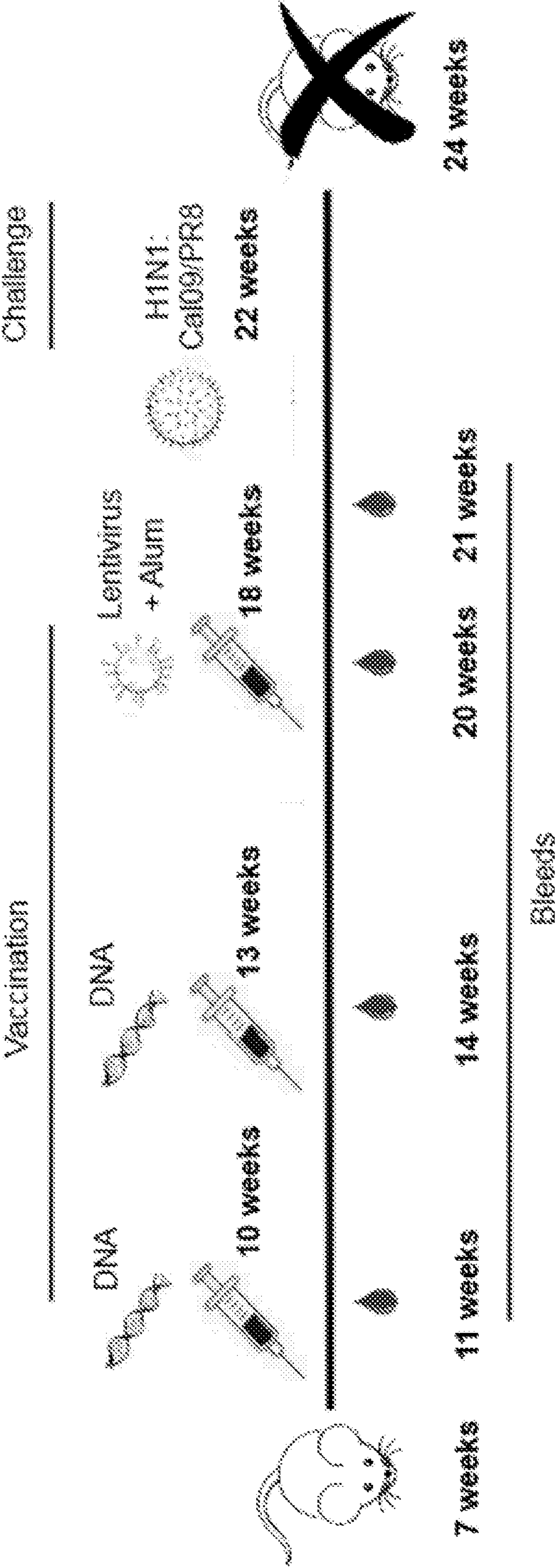
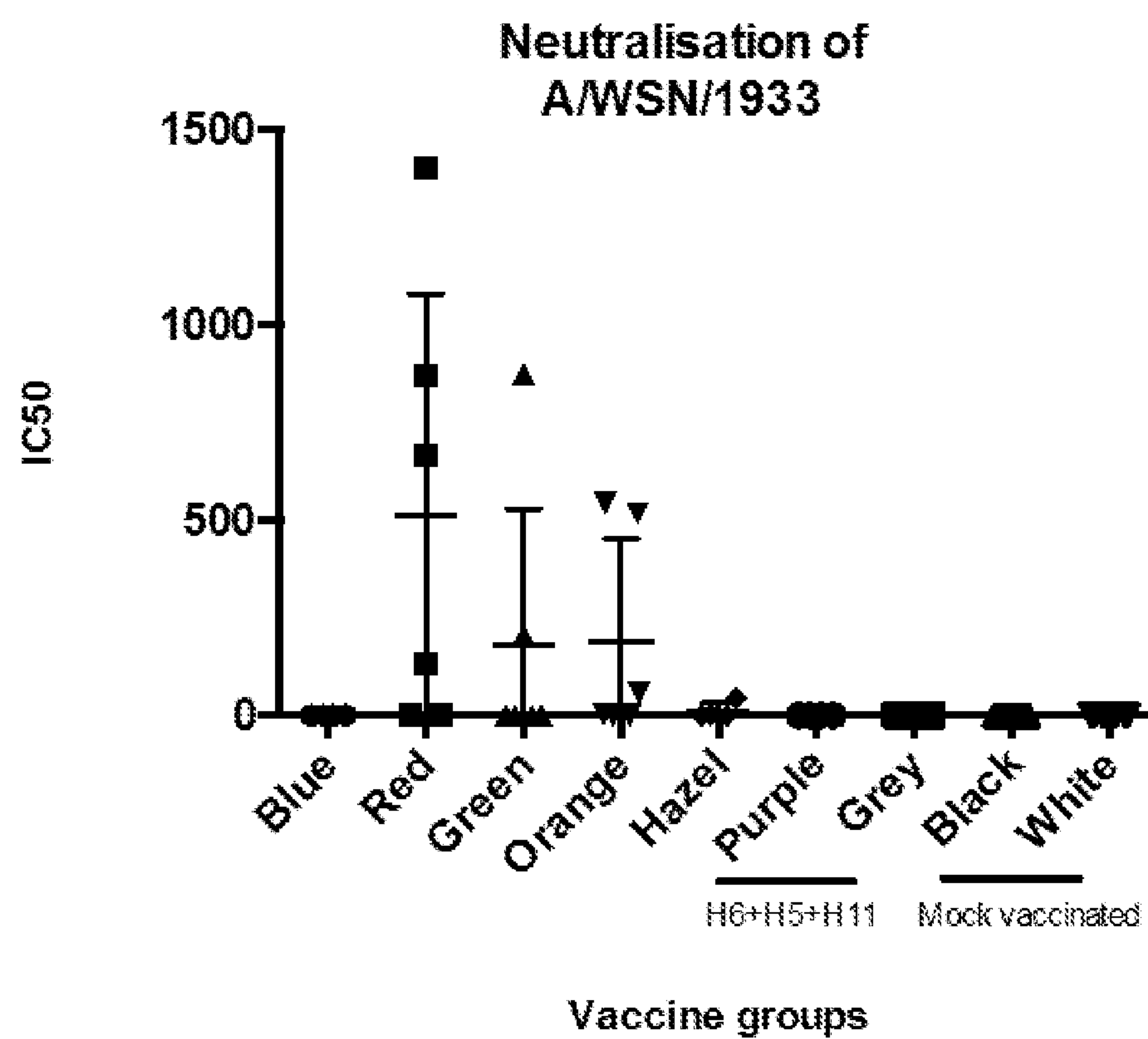


Figure 5C

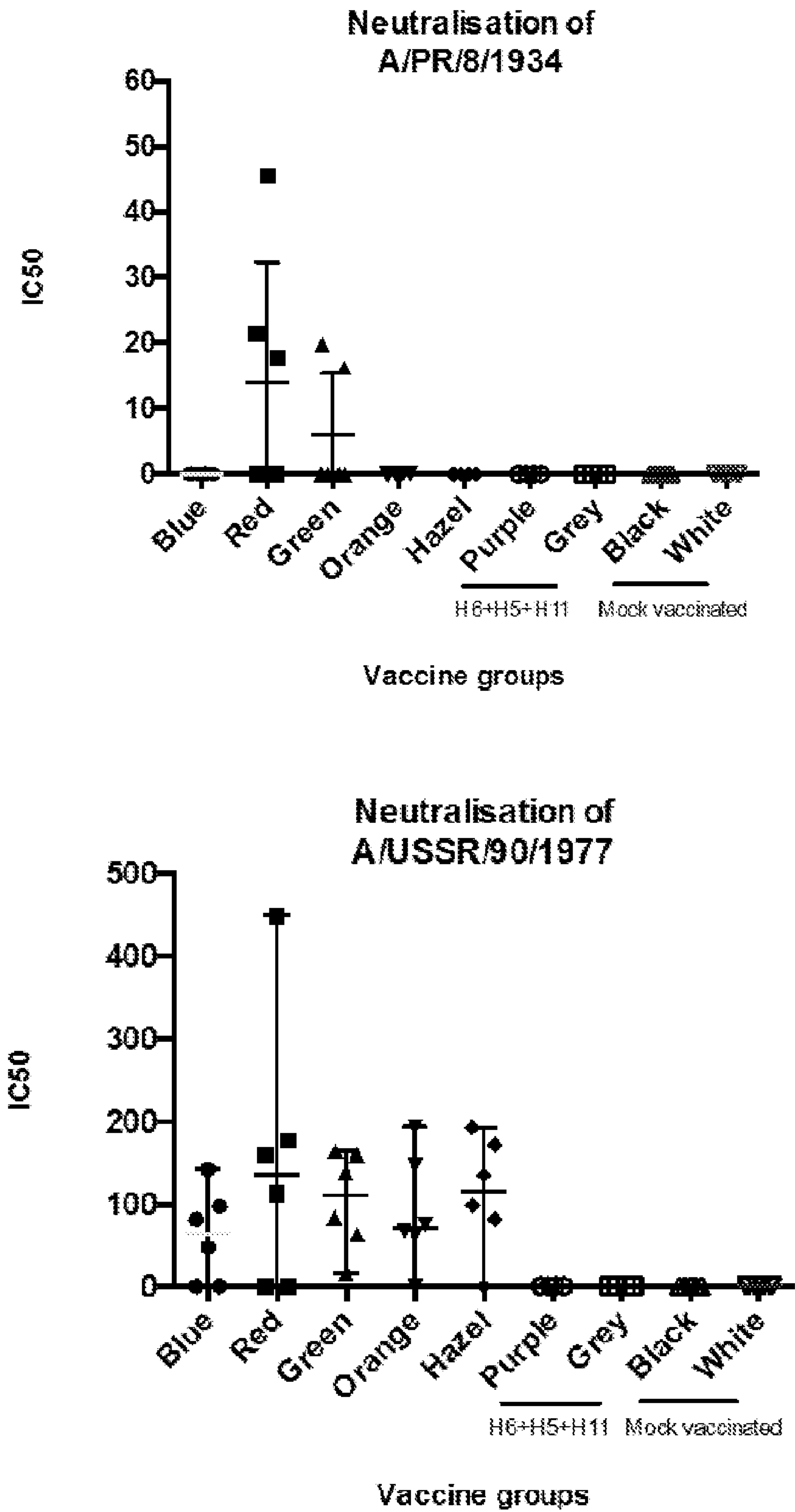


Figure 6A



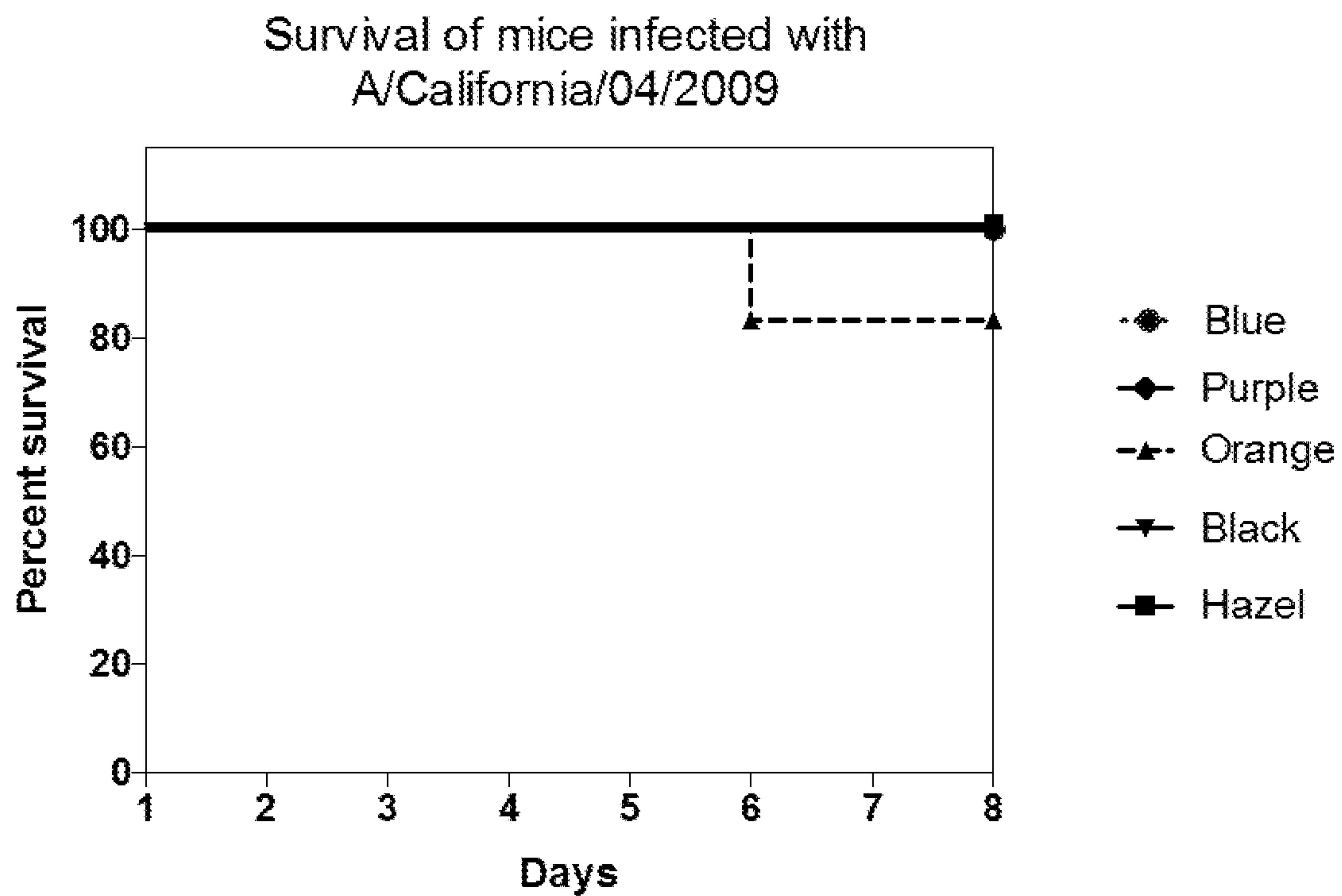
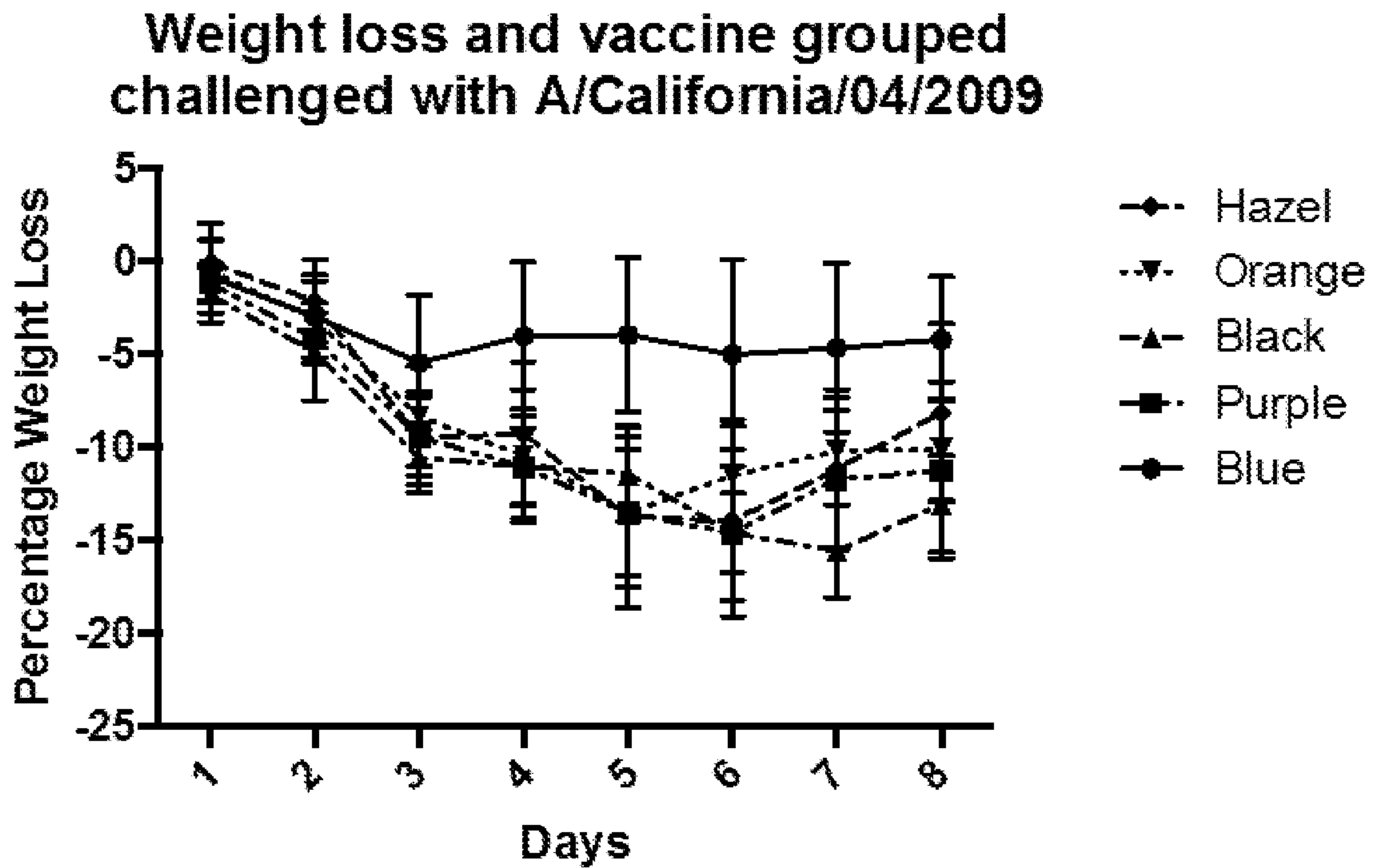


Figures 6D and 6E



- 11 / 29 -

Figures 6G and 6H



- 12 / 29 -

Figures 6I and 6J

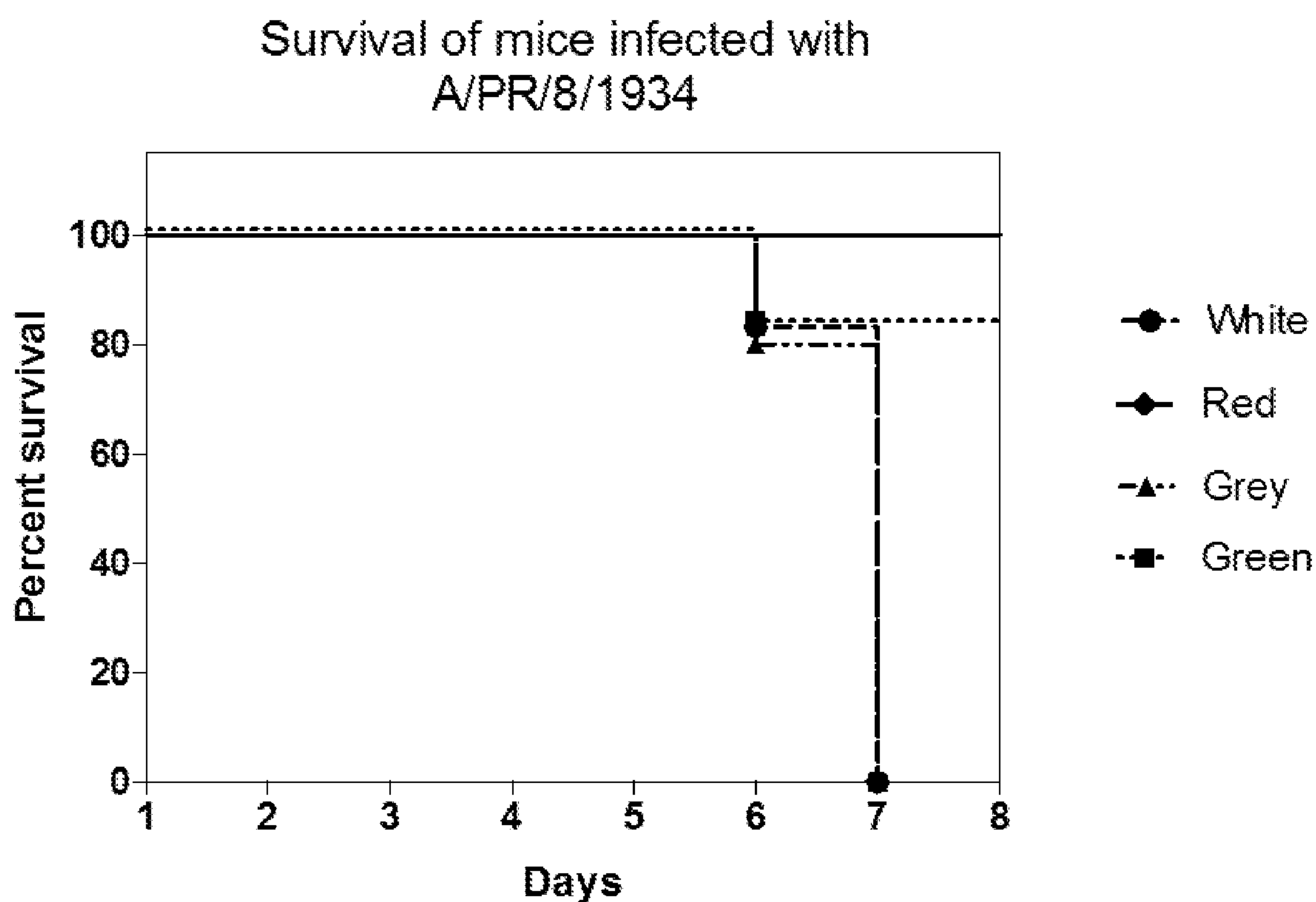
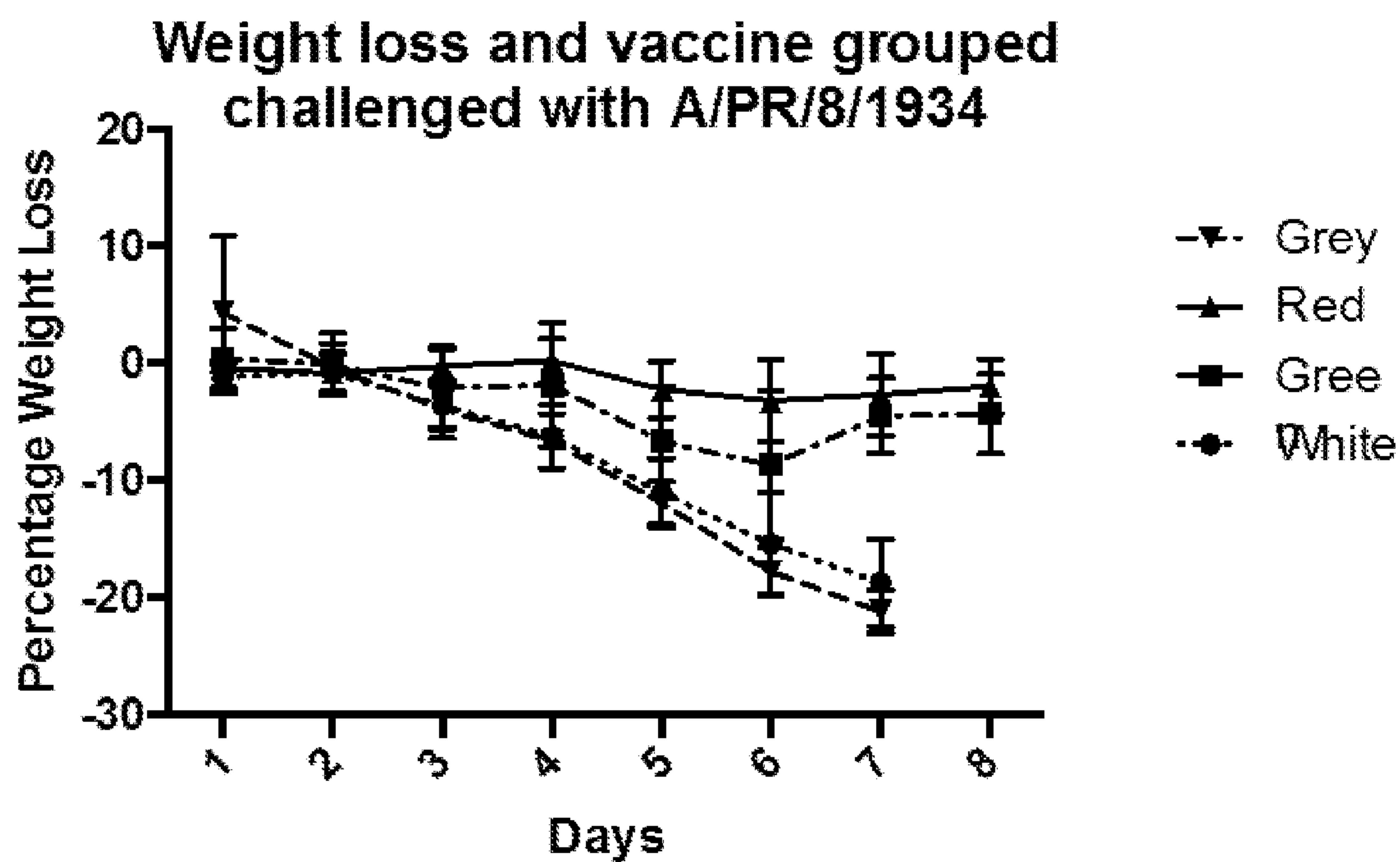


Figure 7A

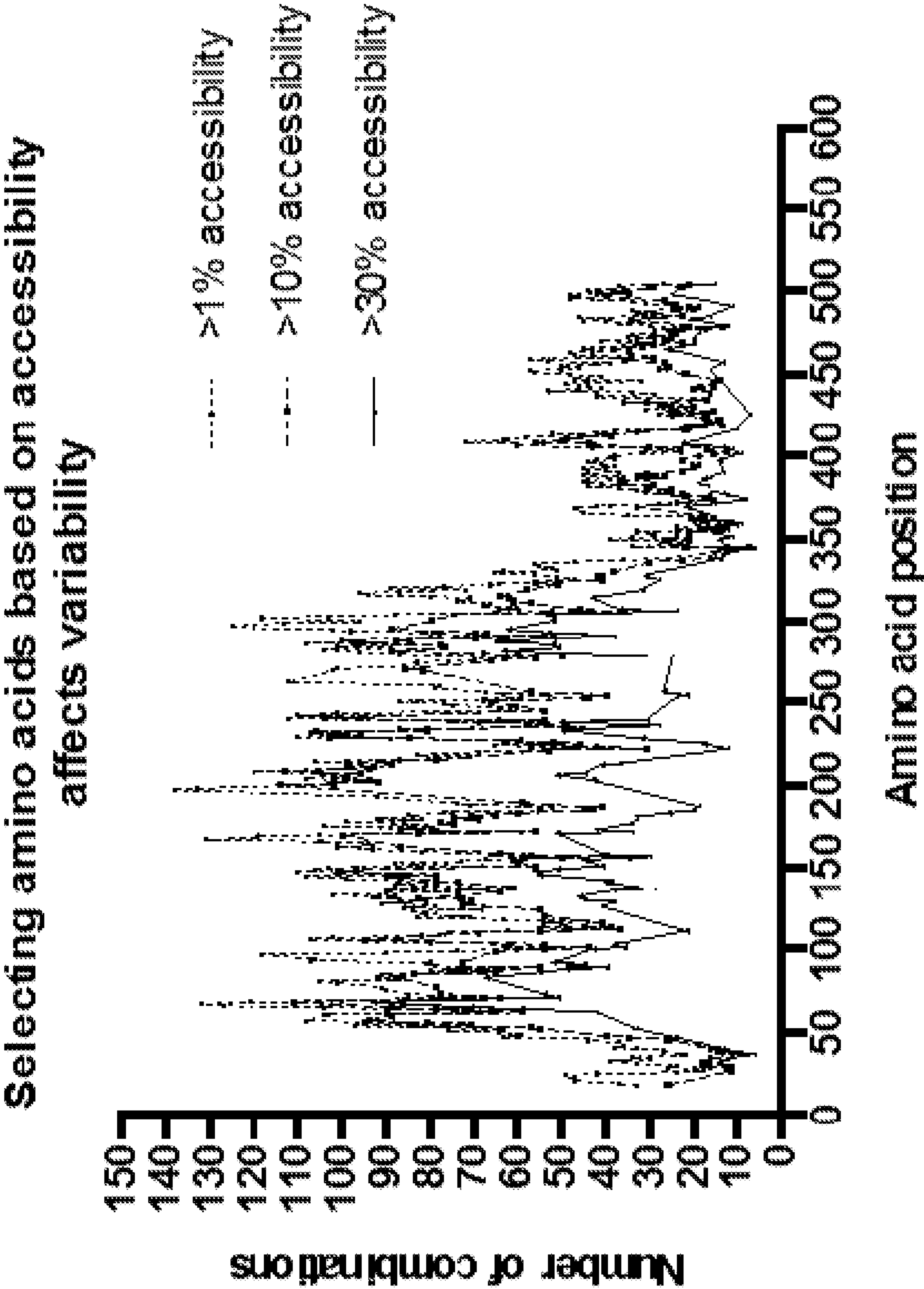


Figure 7B

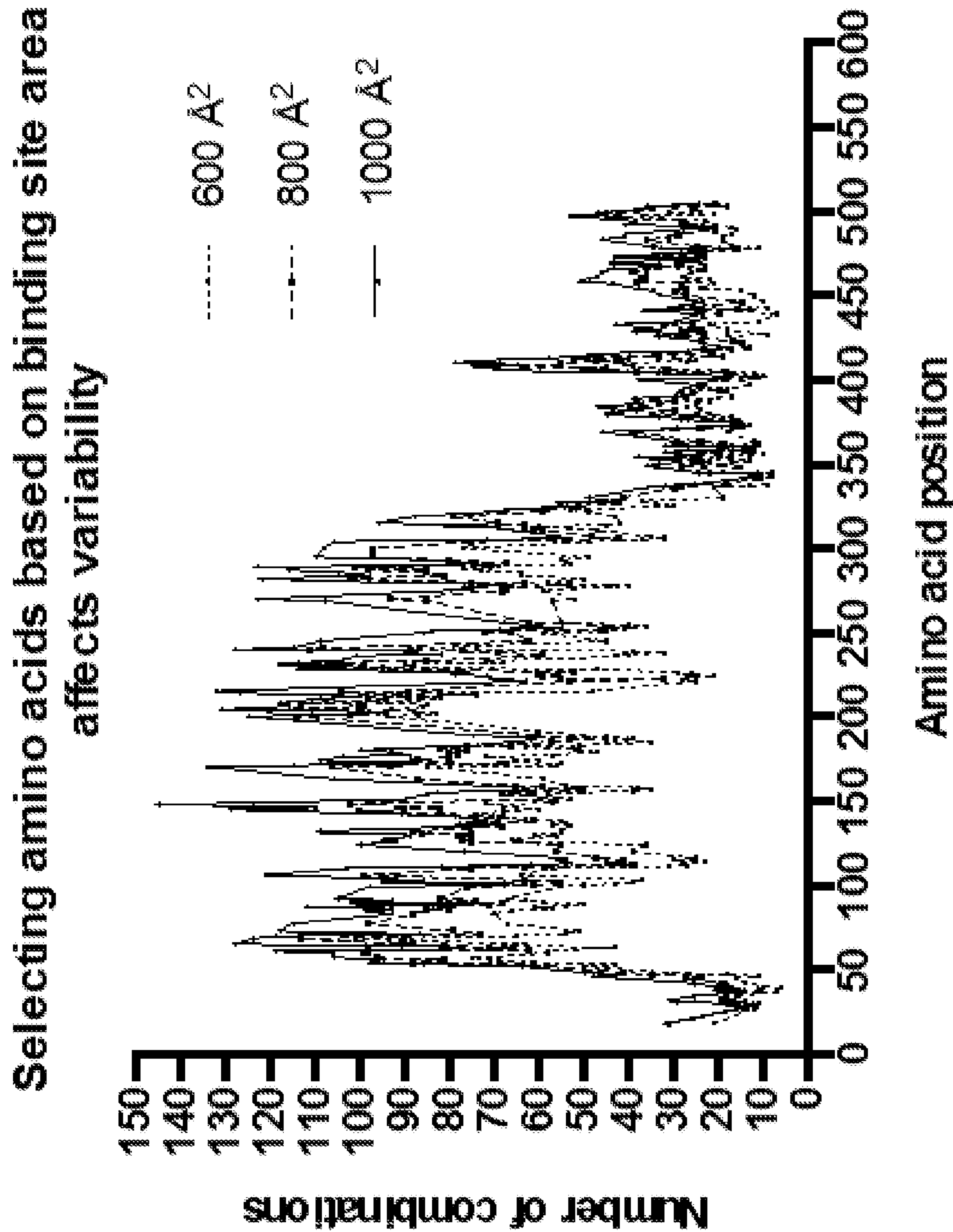
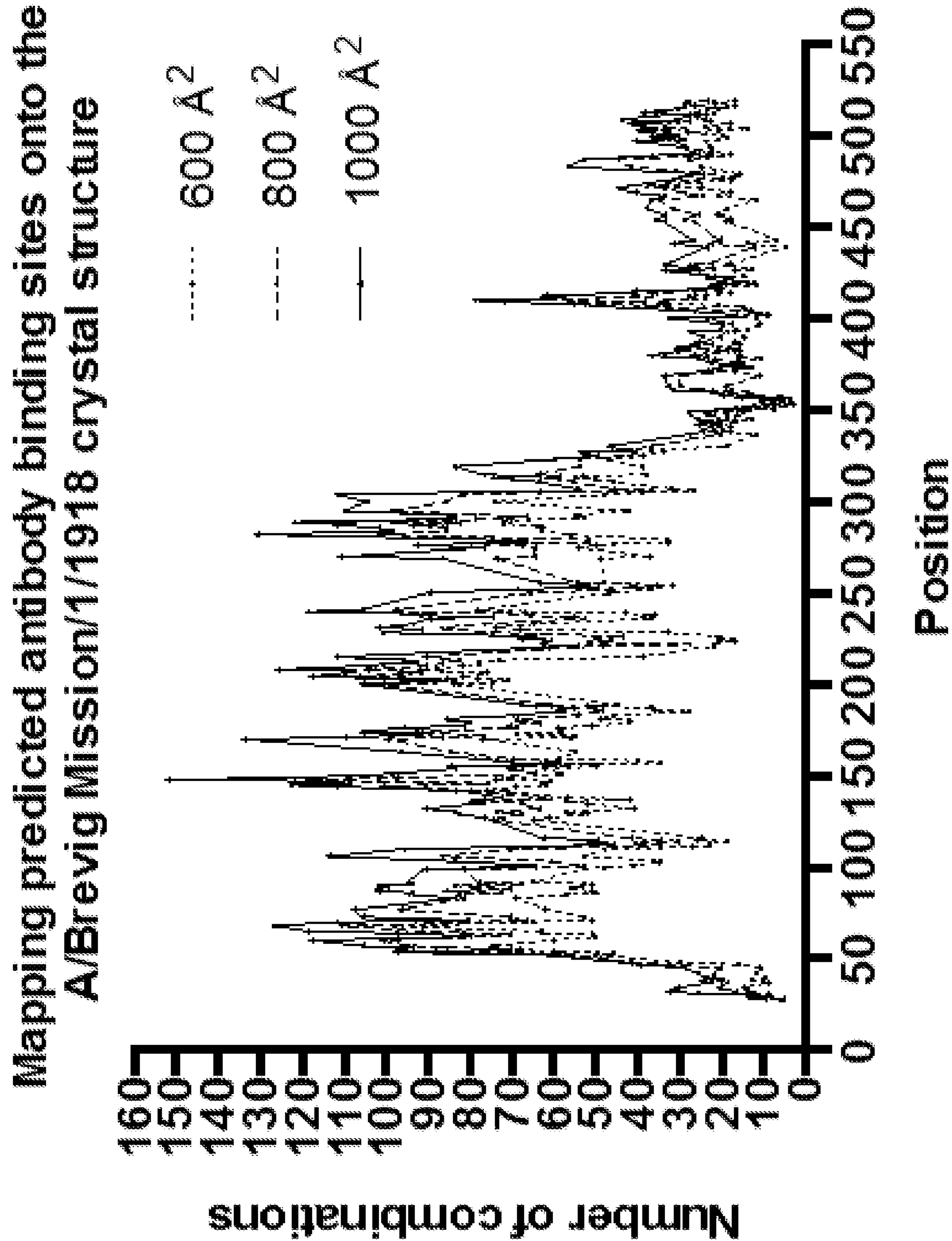


Figure 7C



- 16 / 29 -

Figure 7D

Mapping predicted antibody binding sites onto the
A/Puerto Rico/8/1934 crystal structure

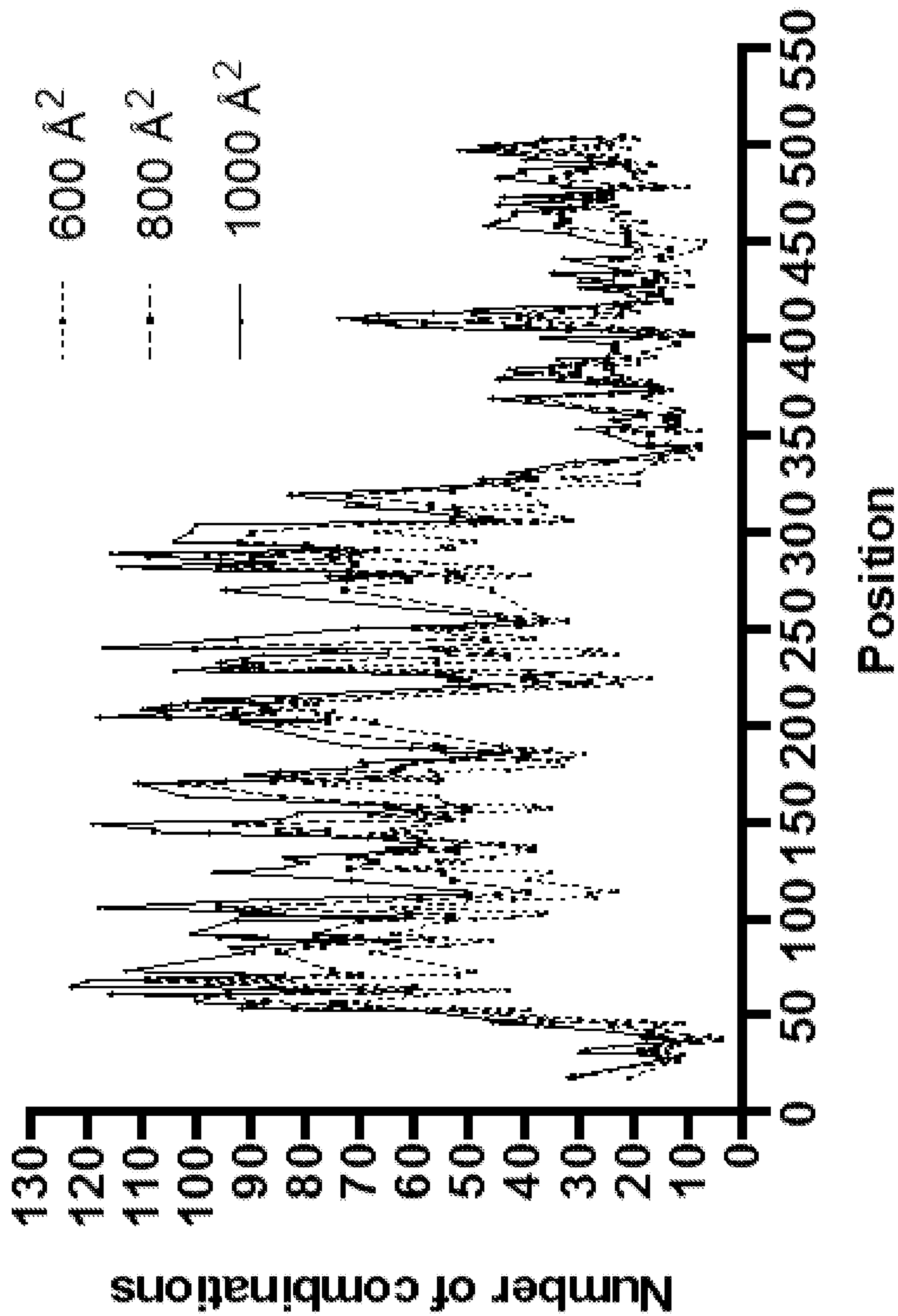


Figure 7E

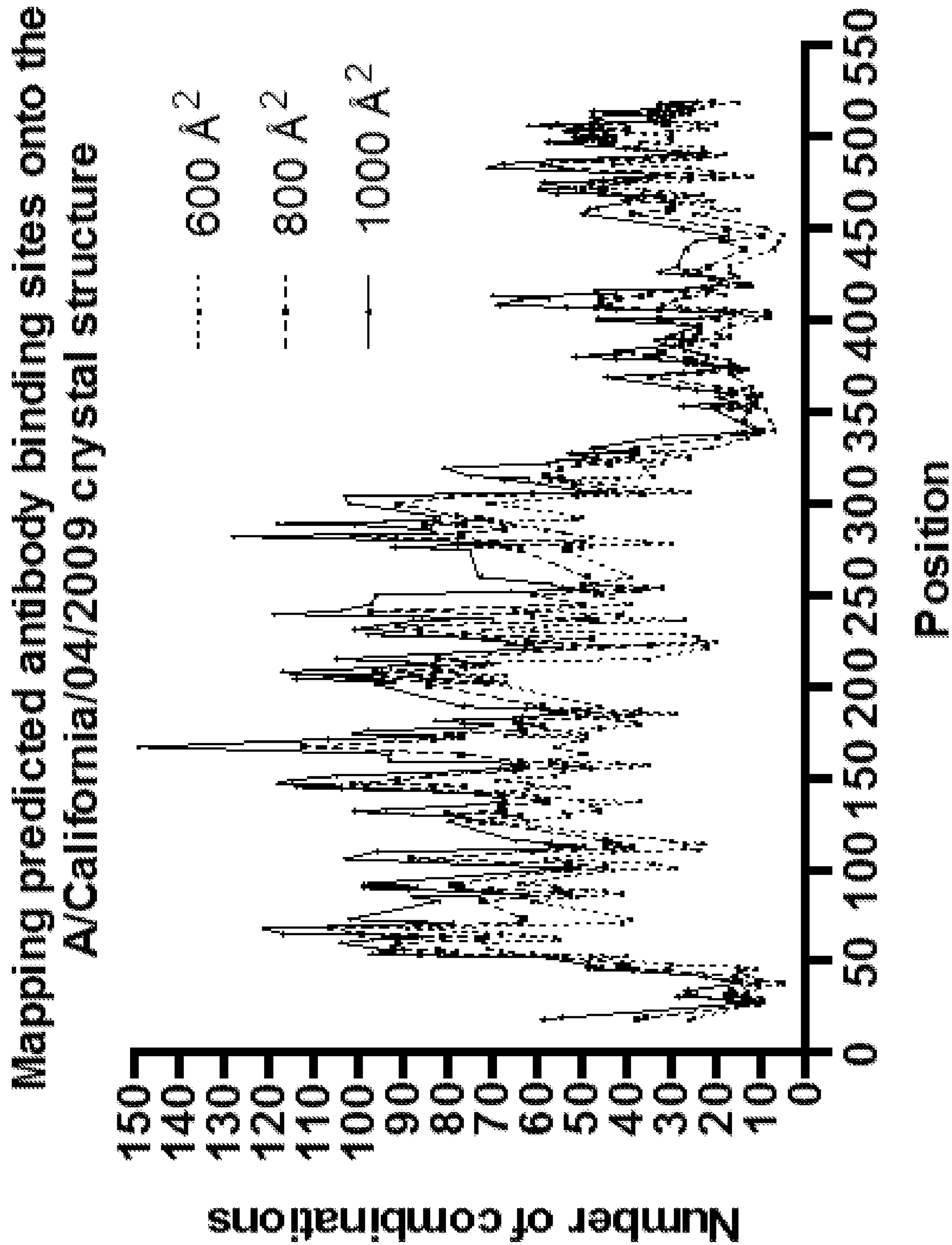


Figure 7F

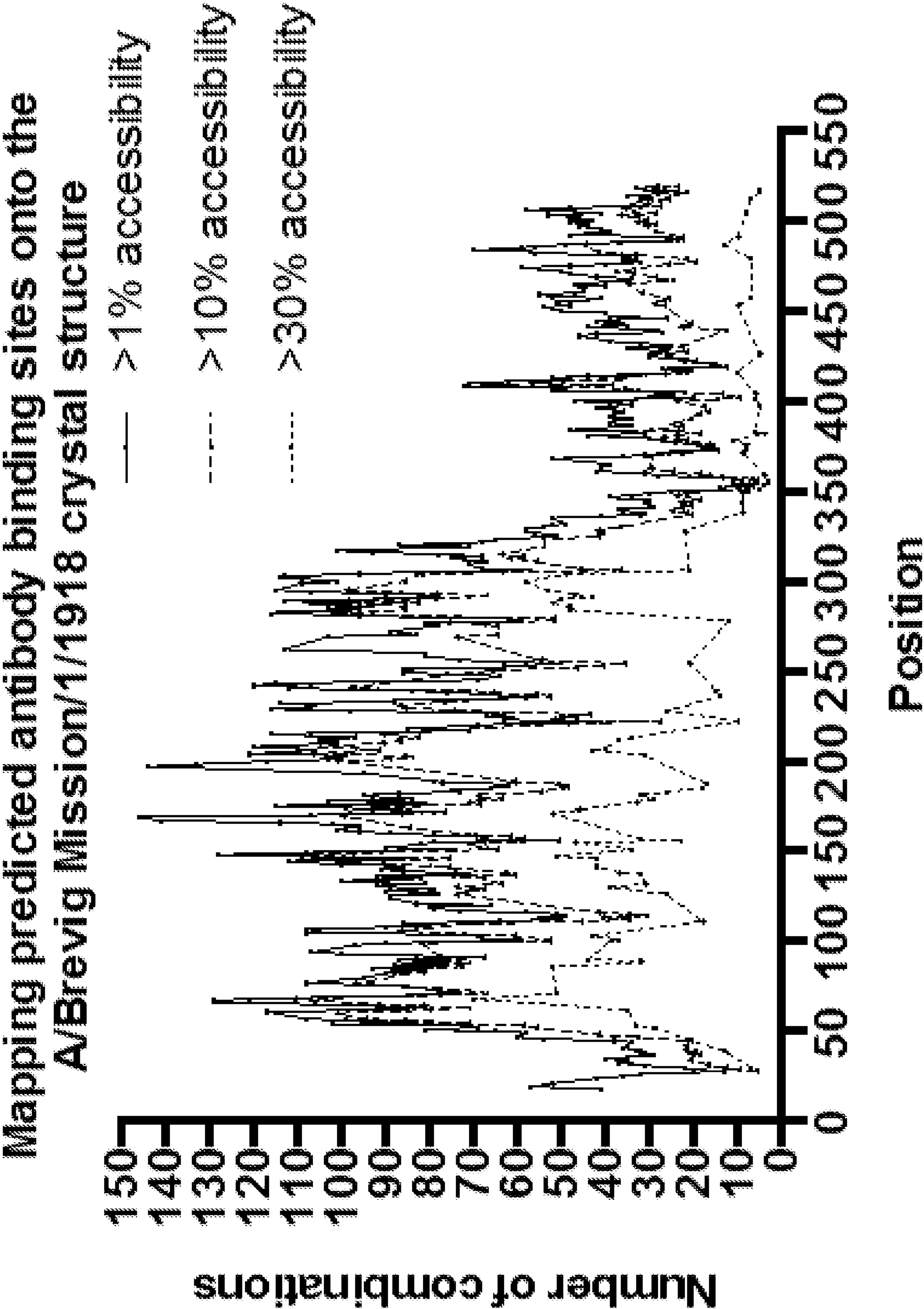
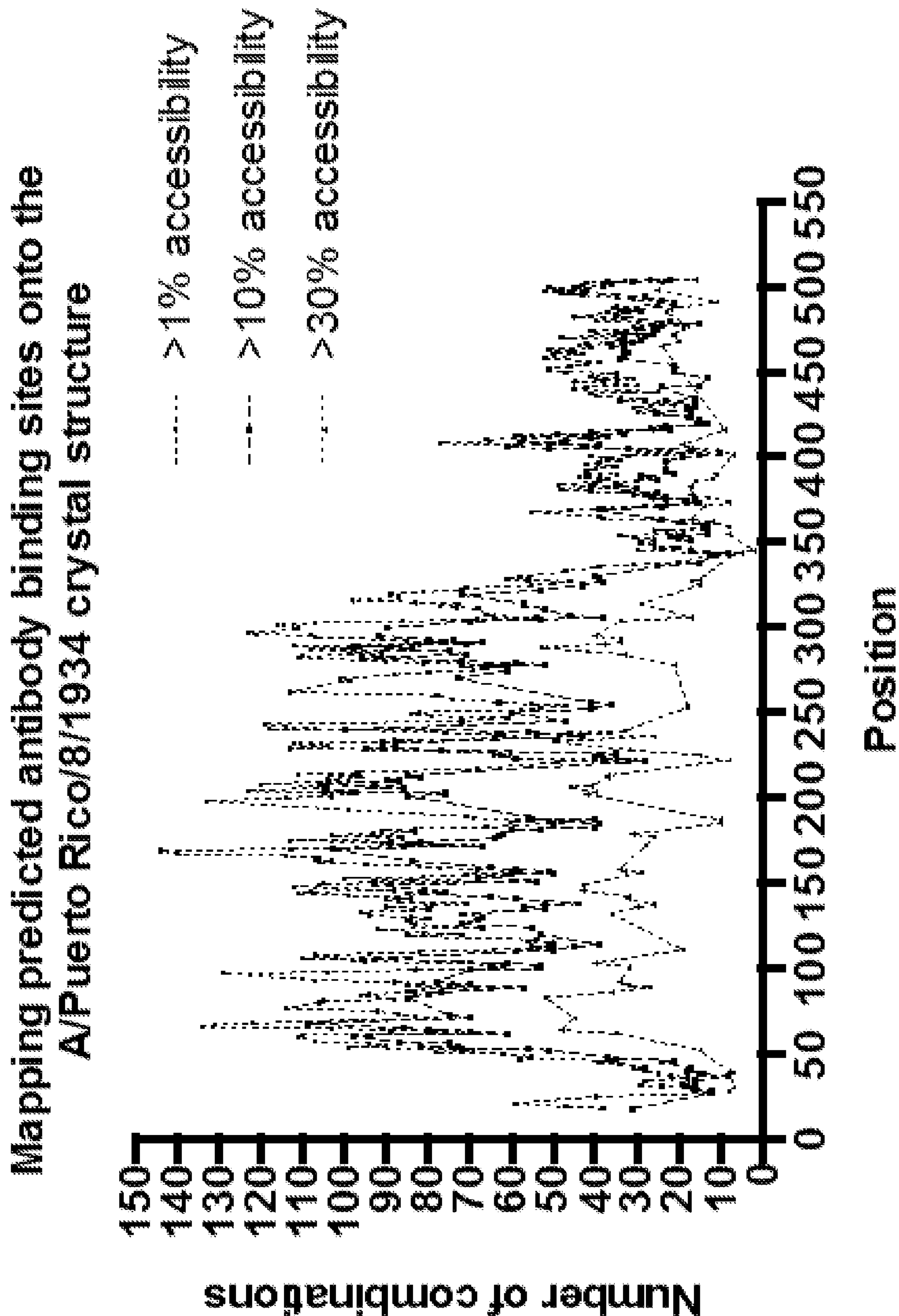
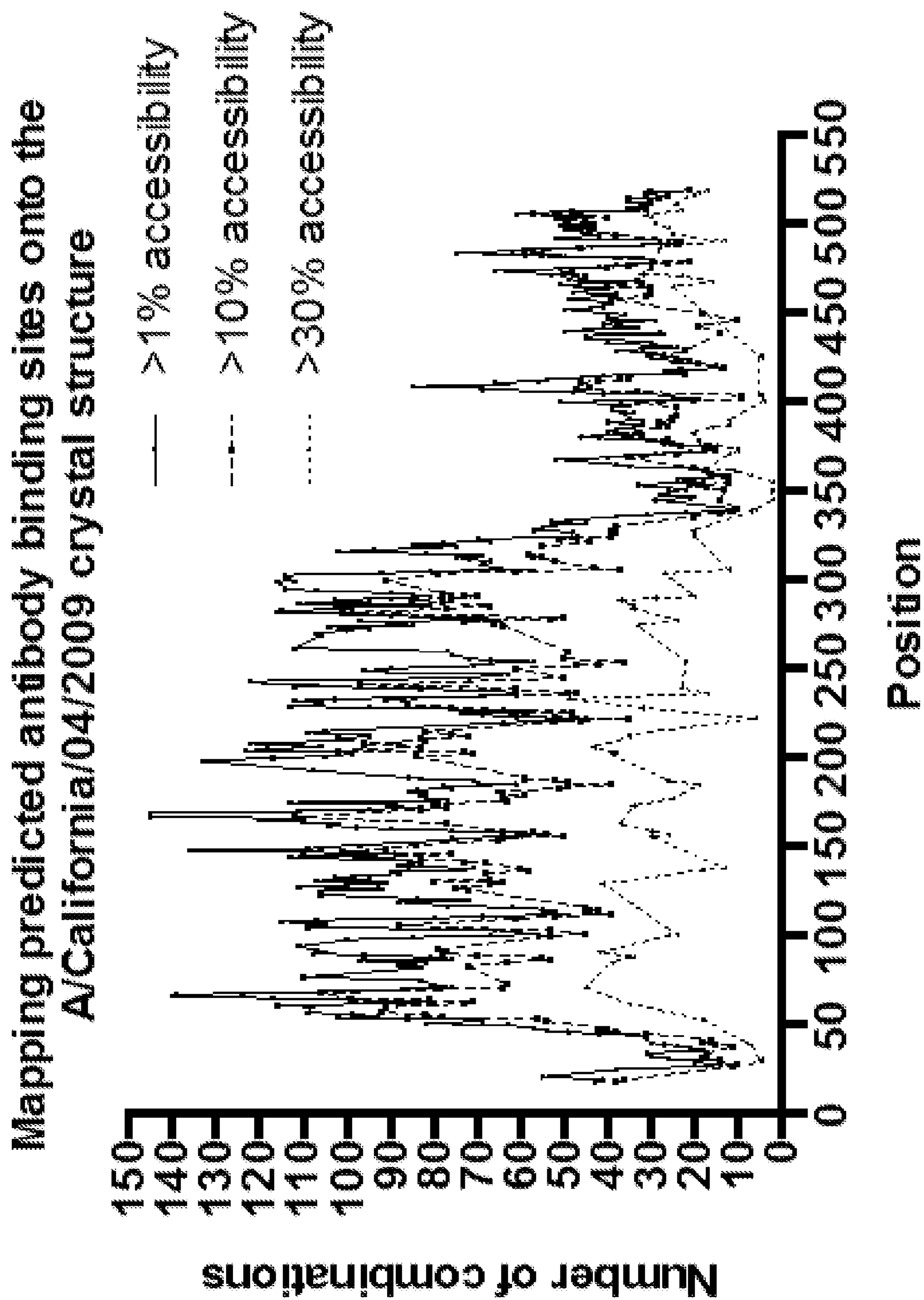


Figure 7G



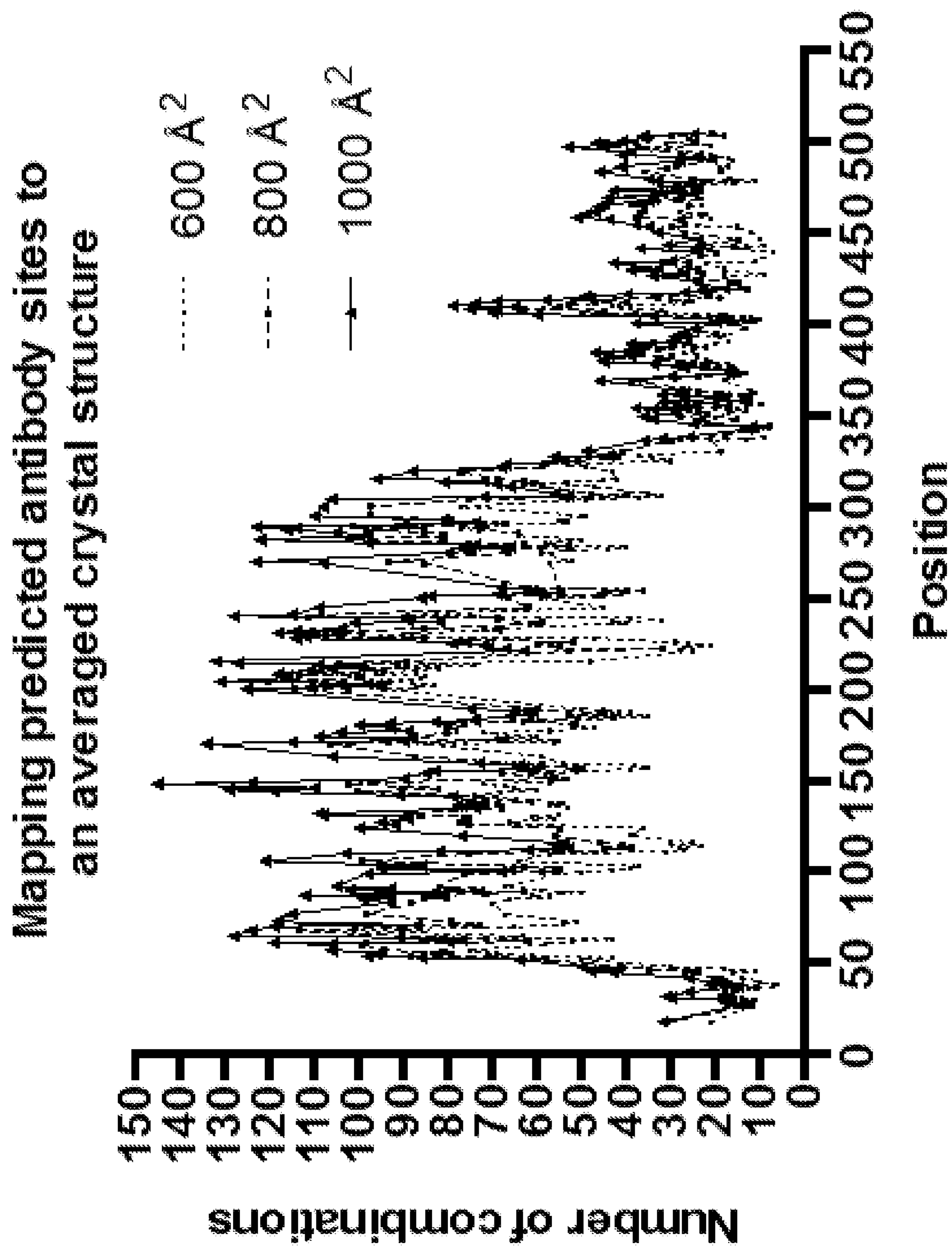
- 20 / 29 -

Figure 7H



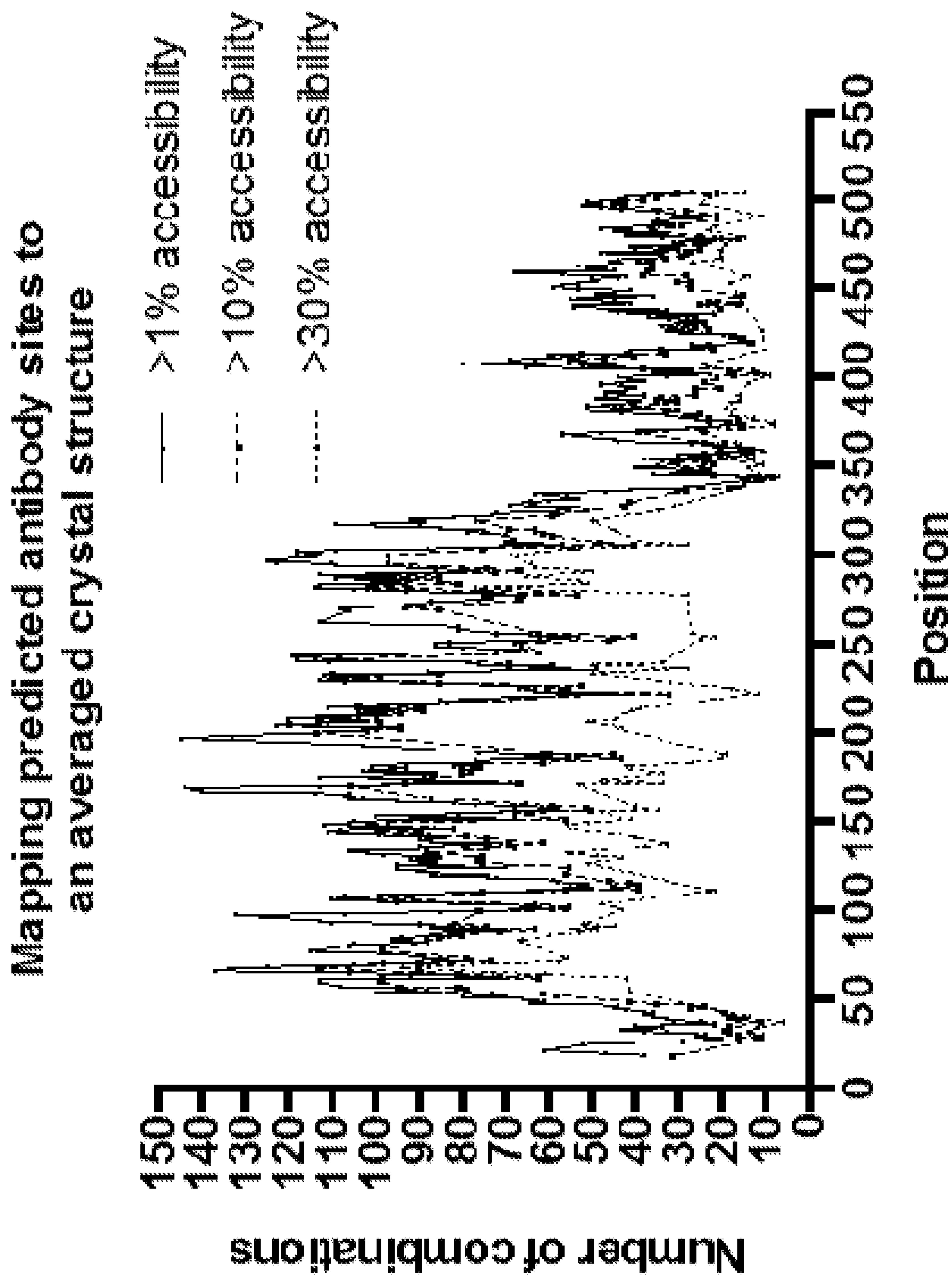
- 21 / 29 -

Figure 7I



- 22 / 29 -

Figure 7J



- 23 / 29 -

Figure 8A

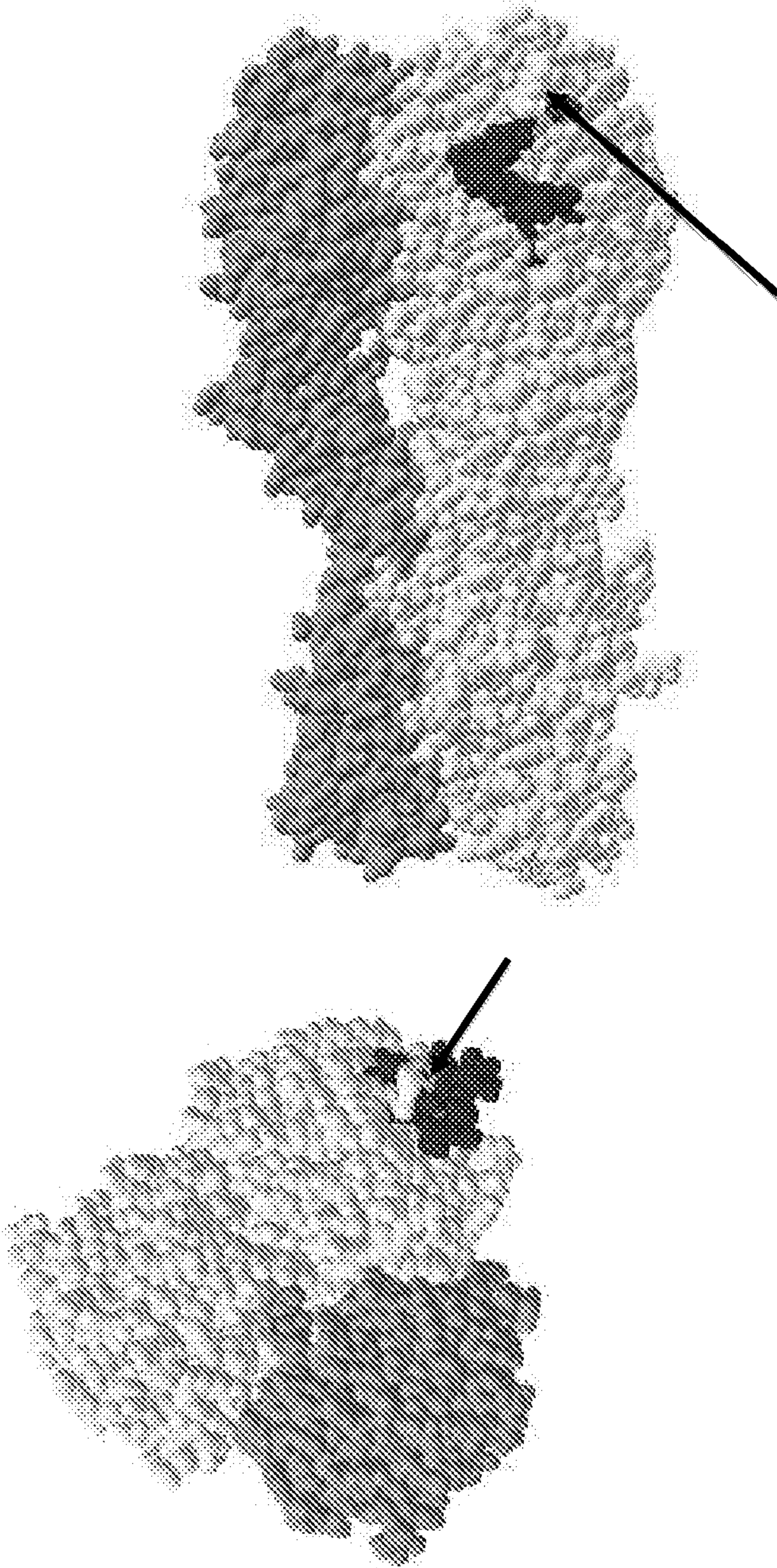


Figure 8B

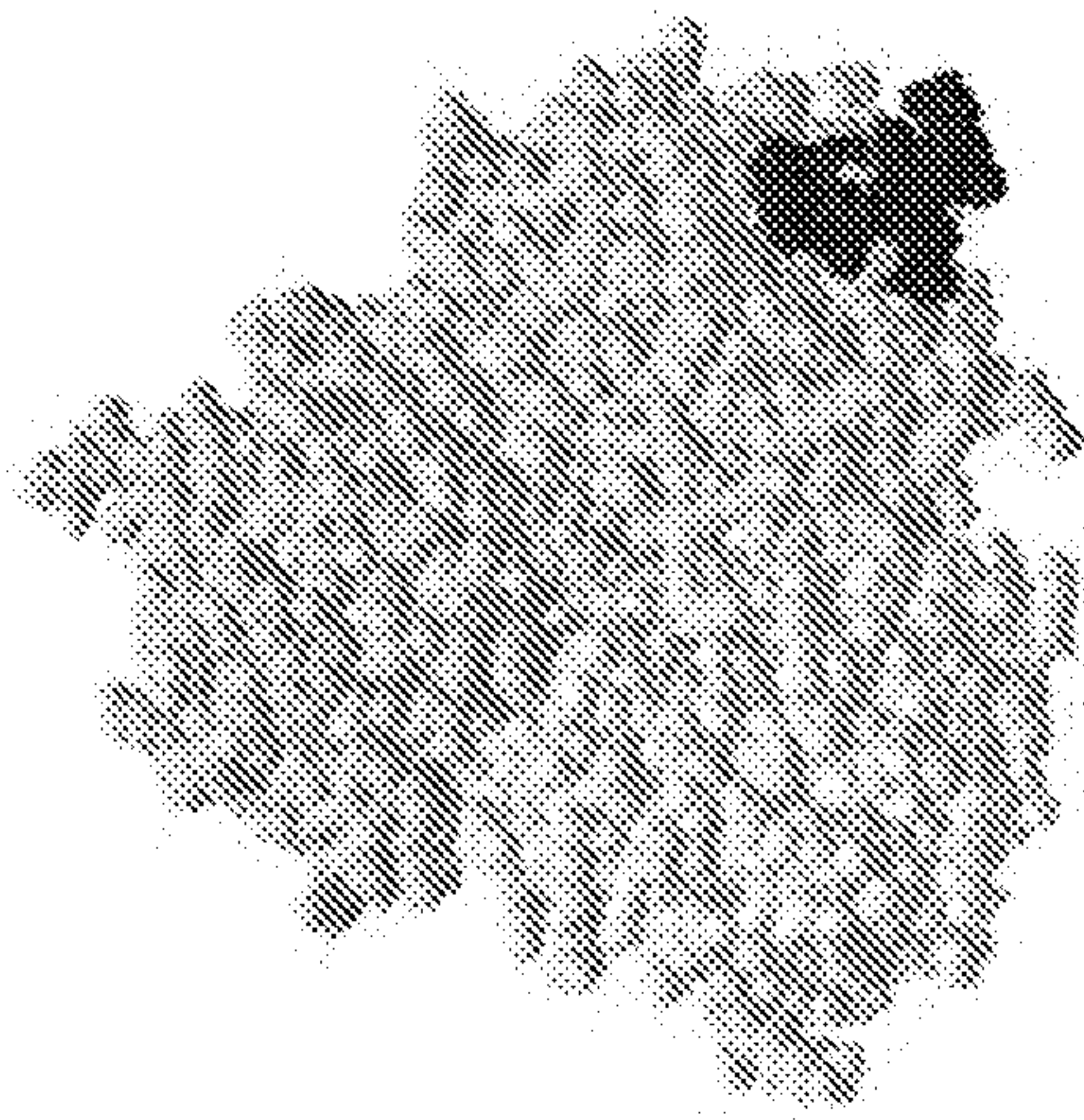
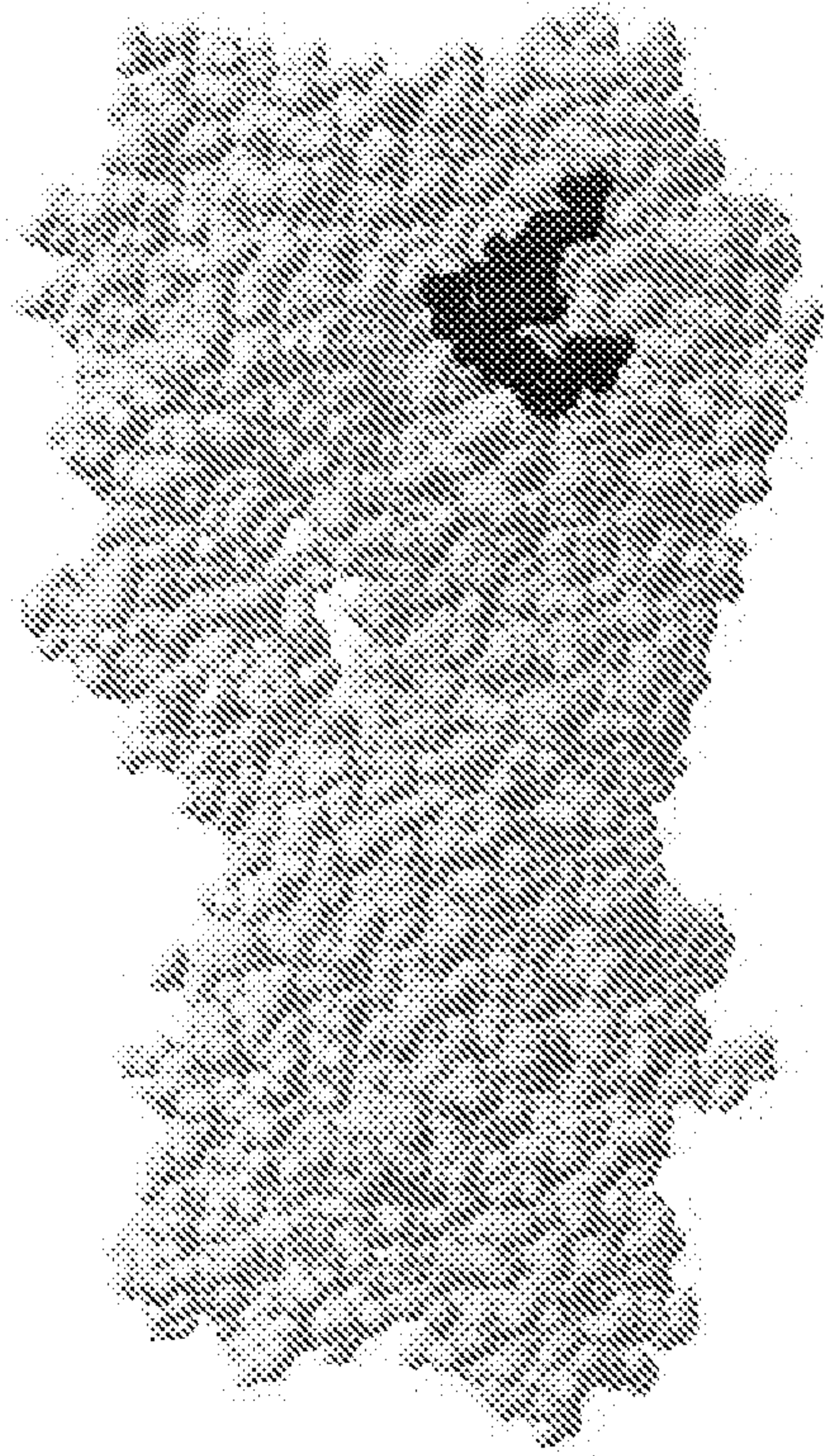


Figure 9A

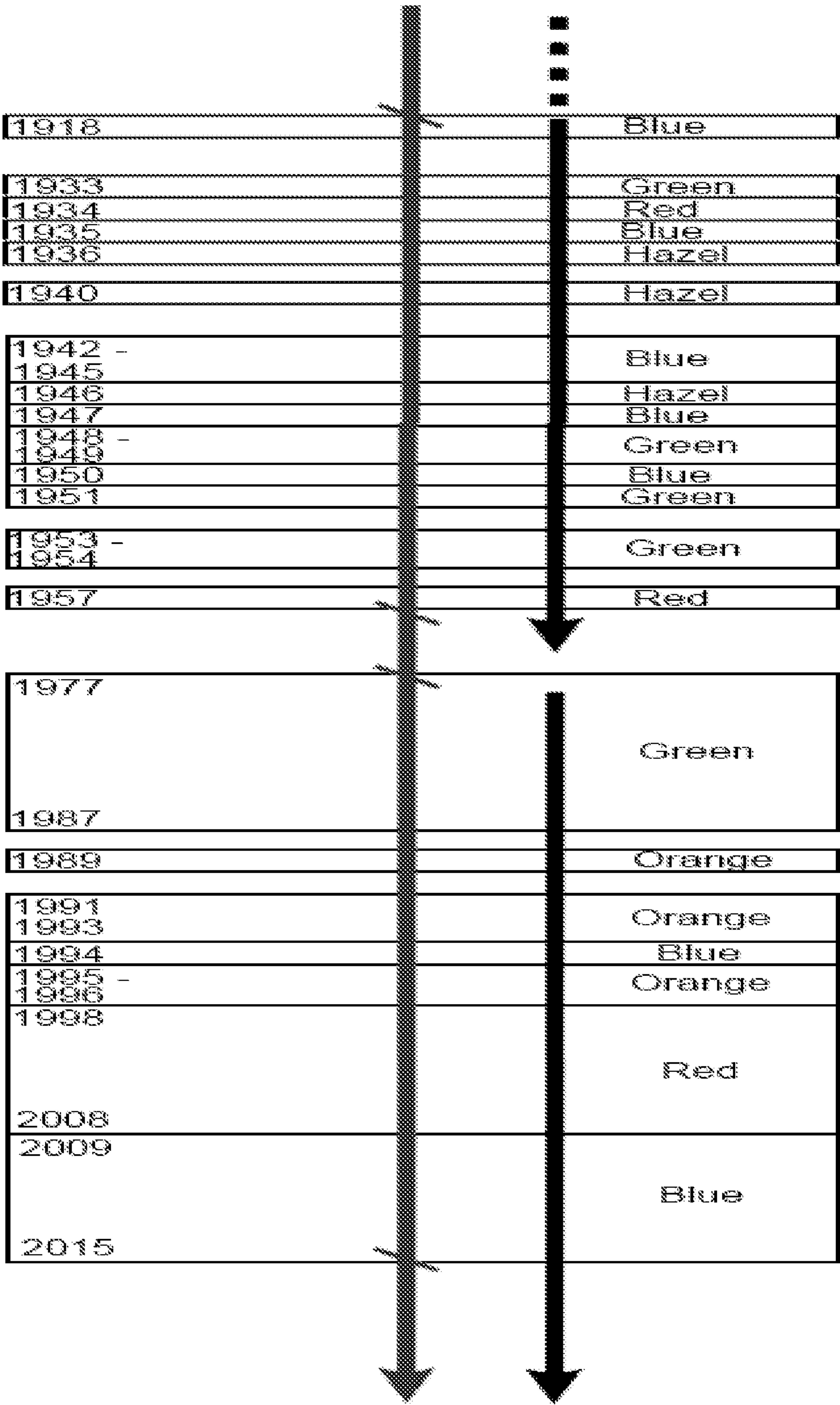


Figure 9B

	83	85	146	147	148	149	151	154	155	156	157	158	159	163	Group 1 (Blue)
2009 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2010 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2011 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2012 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2013 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2014 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
2015 consensus	E	E	N	K	G	V	A	P	H	A	G	A	K	K	
1918 consensus	E	D	T	K	G	V	A	S	H	A	G	A	S	R	
1942 consensus	E	E	T	K	G	V	A	S	H	A	G	K	C	R	
1935 consensus	E	D	T	K	G	V	A	S	H	A	G	K	S	R	
1994 consensus	E	E	N	R	G	V	A	P	H	A	G	A	N	R	
1950 consensus	E	E	I/T	R	G	V	A	S	H	A	G	K	S	K	
1947 consensus	E	E	T	R	G	V	A	S	H	A	G	K	S	K	
1945 consensus	E	E	T	R	G	V	A	S	H	A	G	K	S	R	
1945 consensus	E	E	T	R	G	V	A	S	H	A	G	K	S	R	
1943 consensus	E	E	A	R	G	V	A	S	H	A	G	K	S	R	

	83	E	D/T/N	I	G	V	A	S	H	A	G	K	S	K	Group 2 (Hazel)
1946 consensus	E	E		I	G	V	A	S	H	A	G	K	S	K	
1940 consensus	E	E	N	I	G	V	A	S	H	A	G	K	S	K	
1936 consensus	E	D	N	I	G	V	A	S	H	A	G	K	S	R	

Figure 9C

	Group 3 (Green)											
1951 consensus	E	E	T	R	G	V	A	A	K	K	S	K
1954 consensus	E	E	T	R	G	V	A	A	K	K	S	K
1953 consensus	E	E	T	R	G	V	A	A	R	K	S	K
1948 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1949 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1977 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1978 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1979 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1980 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1982 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1985 consensus	E	E	T	R	G	V	A	K	G	K	S	R
1981 consensus	E	E	T	R	G	V	A	K	G	K	C	R
1933 consensus	E	D	L	K	G	V	A	G/R	G	K	S	R
1983 consensus	E	E	T	K	G	V	A	K	G	K	S	R
1984 consensus	E	E	T	K	G	V	A	K	G	K	S	R
1986 consensus	E	E	T	K	G	V	A	K	G	K	S	R
1987 consensus	E	E	T	K	G	V	A	K	G	K	S	R

Figure 9D

	83	85	146	147	148	149	151	154	155	156	157	158	159	163	Group 4 (Orange)
1989 consensus	E	E	T	K	G	V	A	S	H	N	G	K	S	R	
1992 consensus	E	E	T	K	G	V	A	S	H	N	G	K	S	R	
1993 consensus	E	E	T	K	G	V	A	S	H	N	G	K	S	R	
1991 consensus	K	E	T	K	G	V	A	S	H	N	G	K	S	R	
1995 consensus	E	E	T	K	G	V	A	S	H	N	G	K	S	K	
1996 consensus	E	E	T	K	G	V	A	S	H	N	G	K	S	K	

	83	85	146	147	148	149	151	154	155	156	157	158	159	163	Group 5 (Red)
1998 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
1999 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2000 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2001 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2002 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2003 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2004 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2005 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	R	
2006 consensus	E	E	T	-	G	V	A	S	H	N	G	K	S	K	
2007 consensus	E	E	T	-	G	V	A	S	H	N	G	E	S	R	
2008 consensus	E	E	T	-	G	V	A	S	H	N	G	E	S	R	
1934 consensus	E	D	N	-	G	V	A	S	H	E	G	K	S	R	
1957 consensus	E	E	R	-	G	V	A	P	H	A	R	K	S	K	

Figure 10

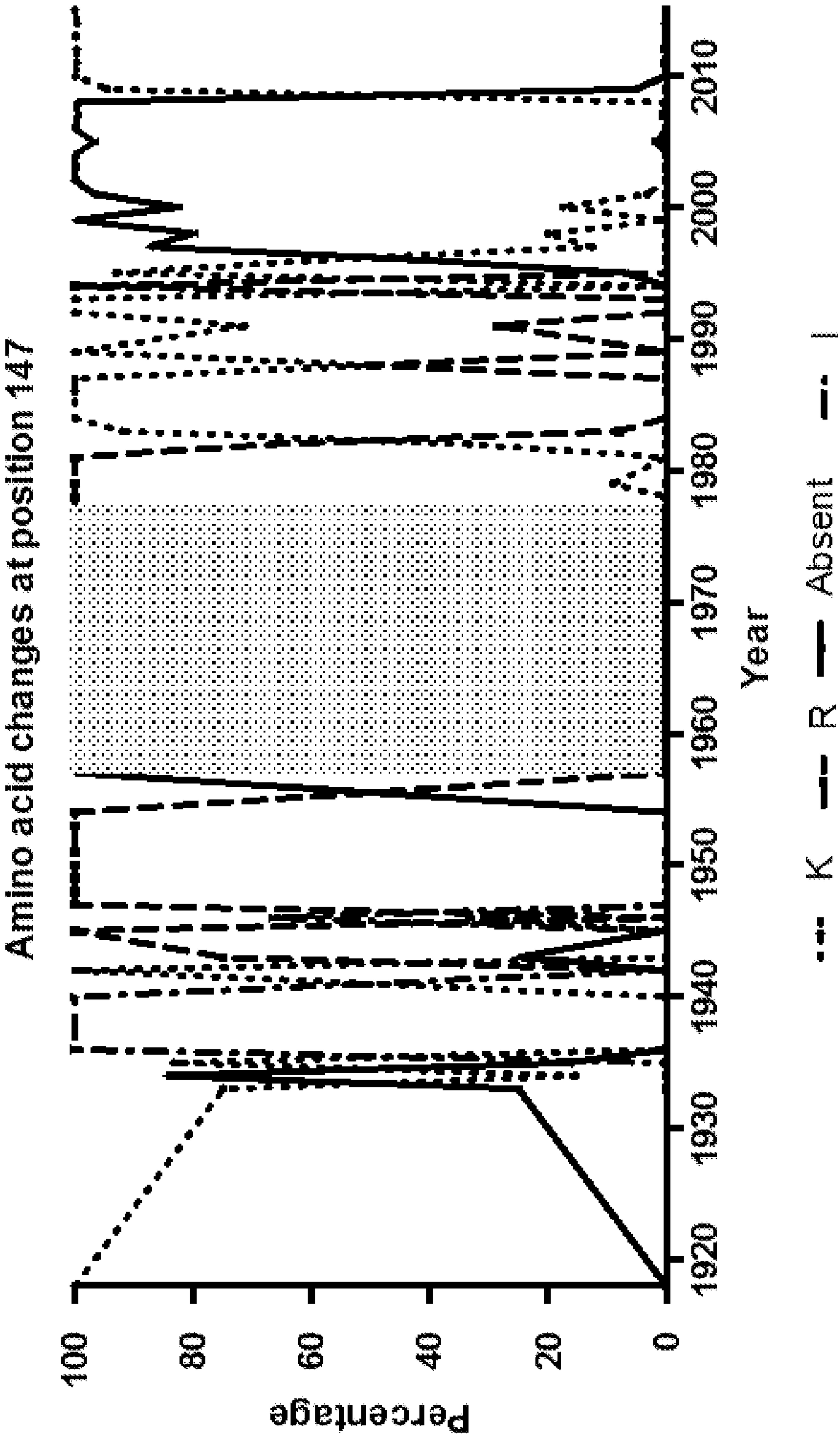


Figure 1

