Identifying immune epitopes sites under selection in influenza hemagglutinin

Austin G. Meyer[1,2,3,*], Claus O. Wilke[1,2]

1 Department of Integrative Biology, Institute for Cellular and Molecular Biology, and Center for Computational Biology and Bioinformatics. The University of Texas at Austin, Austin, TX 78712, USA.

2 Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712, USA.

3 School of Medicine, Texas Tech University Health Sciences Center, Lubbock, TX 79430, USA.

∗ **E-mail: austin.meyer@utexas.edu**

Abstract

Influenza hemagglutinin is among the most studied proteins in all of viral biology. It is both the most variable gene in flu and the protein most responsible for the seasonal re-infection cycle of the human population. There have been dozens of attempts, utilizing as many different methodologies, to identify the sites that are critical for hemagglutinin's seasonal escape from the host immune system. Most of these techniques use some type of sequence analysis to identify sites that are more variable than one would expect from neutral amino acid substitutions; investigators often then assume that highly variable sites are under strong host immune pressure. However, since hemagglutinin is most often analyzed as a test data set for new methodologies in molecular evolution, few investigators try to connect sequence variability to actual immune epitope data. Moreover, in the last decade there has been no attempt to systematically re-analyze flu despite a ten-fold growth in available data and the crystallization of well-established molecular evolutionary techniques. Further complicating matters, there are a surprisingly large number of technical complexities necessary to appropriately draw conclusions about flu biology. As a result, a huge number of available analyses belie a dearth of analyses that are useful for understanding flu evolution. For hemagglutinin H3, we have re-analyzed all currently available sequences and curated all experimental immune epitope data. We find that epitope sites are enriched for sites under positive selection. In addition, we find there are a large number of sites that are under diversifying selection that have no experimental justification for being under immune pressure; likewise there are a large number of epitope sites that are not under diversifying selection.

Author Summary

Introduction

Materials and Methods

Results

Subsection 1

SubSubsection 1.1

Subsection 2

Discussion

Acknowledgments

References

Figure Legends

Tables

Gene Numbering	Protein Numbering	Wiley 1981	Wiley 1987	Bush 1999	Notes
1	NA	-	-	-	
2	NA	-	-	-	
3	NA	-	-	-	
4	NA	-	-	-	
5	NA	-	-	-	
6	NA	-	-	-	
7	NA	-	-	-	
8	NA	-	-	-	
9	NA	-	-	-	
10	NA	-	-	-	
11	NA	-	-	-	
12	NA	-	-	-	
13	NA	-	-	-	
14	NA	-	-	-	
15	NA	-	-	-	
16	NA	-	-	-	
17	1	-	-	-	
18	2	-	-	-	
19	3	N	N	-	
20	4	-	-	-	
21	5	-	-	-	
22	6	-	-	-	

23	7	-	-	-
24	8	-	-	-
25	9	N	N	-
26	10	-	-	-
27	11	-	-	-
28	12	-	-	-
29	13	-	-	-
30	14	-	-	-
31	15	-	-	-
32	16	-	-	-
33	17	-	-	-
34	18	-	-	-
35	19	-	-	-
36	20	-	-	-
37	21	-	-	-
38	22	-	-	-
39	23	-	-	-
40	24	-	-	-
41	25	-	-	-
42	26	-	-	-
43	27	-	-	-
44	28	-	-	-
45	29	-	-	-
46	30	-	-	-
47	31	N	N	-
48	32	-	-	-
49	33	-	-	-
50	34	-	N	-
51	35	-	-	-
52	36	-	-	-
53	37	-	-	-
54	38	-	-	-
55	39	-	-	-
56	40	-	-	-
57	41	-	-	-
58	42	-	-	-
59	43	-	-	-
60	44	-	-	C
61	45	-	-	C
62	46	-	-	C
63	47	-	-	C
64	48	-	C	C
65	49	-	-	-
66	50	-	C	C
67	51	-	-	C
68	52	-	-	-
69	53	C	C	C
70	54	C	C	C
71	55	-	-	-

72	56	-	-	-
73	57	-	-	E
74	58	-	N	-
75	59	-	-	E
76	60	-	-	-
77	61	-	-	-
78	62	-	E	E
79	63	N	E	E
80	64	-	-	-
81	65	-	-	-
82	66	-	-	-
83	67	-	N	E
84	68	-	-	-
85	69	-	-	-
86	70	-	-	-
87	71	-	-	-
88	72	-	-	-
89	73	-	-	-
90	74	-	-	-
91	75	-	-	E
92	76	-	-	-
93	77	-	-	-
94	78	N	E	E
95	79	-	-	-
96	80	-	-	E
97	81	-	E	E
98	82	-	-	E
99	83	N	E	E
100	84	-	-	-
101	85	-	-	-
102	86	-	-	E
103	87	-	-	E
104	88	-	-	E
105	89	-	-	-
106	90	-	-	-
107	91	-	C	E
108	92	-	C	E
109	93	-	-	-
110	94	-	-	E
111	95	-	-	-
112	96	-	-	D
113	97	-	-	-
114	98	-	-	-
115	99	-	-	-
116	100	-	-	-
117	101	-	-	-
118	102	-	-	D
119	103	-	-	D
120	104	-	-	-

121	105	-	-	-	
122	106	-	-	-	
123	107	-	-	-	
124	108	-	-	-	
125	109	-	-	E	
126	110	N	-	-	
127	111	-	-	-	
128	112	-	-	-	
129	113	-	-	-	
130	114	-	-	-	
131	115	-	-	-	
132	116	-	-	-	
133	117	-	-	D	
134	118	-	-	-	
135	119	-	-	-	
136	120	-	-	-	
137	121	-	-	D	
138	122	A	C	A	Probably wrong in 1987
139	123	-	-	-	
140	124	-	N	A	
141	125	-	-	-	
142	126	N	B	A	Probably wrong in 1987
143	127	-	-	-	
144	128	-	M	B	
145	129	-	B	B	
146	130	-	-	A	
147	131	-	-	A	
148	132	-	A	A	
149	133	A	A	A	
150	134	-	-	-	
151	135	-	M	A	
152	136	-	-	-	
153	137	A	A	A	
154	138	-	-	A	
155	139	-	-	-	
156	140	-	-	A	
157	141	-	-	-	
158	142	-	-	A	
159	143	A	A	A	
160	144	A	A	A	
161	145	A	A	A	
162	146	A	A	A	
163	147	-	-	-	
164	148	-	-	-	
165	149	-	-	-	
166	150	-	-	A	
167	151	-	-	-	
168	152	-	-	A	
169	153	-	-	-	

170	154	-	-	-
171	155	B	B	B
172	156	-	B	B
173	157	-	N	B
174	158	-	B	B
175	159	-	B	B
176	160	-	B	B
177	161	-	-	-
178	162	-	-	-
179	163	-	B	B
180	164	N	N	B
181	165	-	-	B
182	166	-	-	-
183	167	-	-	D
184	168	-	-	A
185	169	-	-	-
186	170	-	-	D
187	171	-	-	D
188	172	-	D	D
189	173	-	D	D
190	174	N	D	D
191	175	-	-	D
192	176	-	-	D
193	177	-	-	D
194	178	-	-	-
195	179	-	-	D
196	180	-	-	-
197	181	-	-	-
198	182	N	N	D
199	183	-	-	-
200	184	-	-	-
201	185	-	-	-
202	186	B	-	B
203	187	-	-	B
204	188	B	B	B
205	189	B	B	B
206	190	-	-	B
207	191	-	N	-
208	192	-	-	B
209	193	B	B	B
210	194	-	-	B
211	195	-	-	-
212	196	-	-	B
213	197	-	B	B
214	198	-	B	B
215	199	-	M	-
216	200	-	-	-
217	201	D	D	D
218	202	-	-	-

219	203	-	-	D	
220	204	-	-	-	
221	205	D	M	-	
222	206	-	-	-	
223	207	D	D	D	
224	208	D	N	D	Not plotted in 1987
225	209	-	-	D	
226	210	-	-	-	
227	211	-	-	-	
228	212	-	-	D	
229	213	-	N	D	Not plotted in 1987
230	214	-	-	D	
231	215	-	-	D	
232	216	-	-	D	
233	217	D	D	D	
234	218	-	M	D	
235	219	-	M	D	
236	220	D	-	-	
237	221	-	-	-	
238	222	-	-	-	
239	223	-	-	-	
240	224	-	-	-	
241	225	-	-	-	
242	226	N	D	D	
243	227	-	-	D	
244	228	N	N	D	
245	229	-	-	D	Illegible in 1987
246	230	-	-	D	
247	231	-	-	-	
248	232	-	-	-	
249	233	-	-	-	
250	234	-	-	-	
251	235	-	-	-	
252	236	-	-	-	
253	237	-	-	-	
254	238	-	-	D	
255	239	-	-	-	
256	240	-	-	D	
257	241	-	-	-	
258	242	N	D	D	
259	243	-	-	-	
260	244	-	D	D	
261	245	-	-	-	
262	246	-	D	D	
263	247	-	-	D	
264	248	-	-	D	
265	249	-	-	-	
266	250	-	-	-	
267	251	-	-	-	

268	252	-	-	-
269	253	-	-	-
270	254	-	-	-
271	255	-	-	-
272	256	-	-	-
273	257	-	-	-
274	258	-	-	-
275	259	-	-	-
276	260	N	N	E
277	261	-	N	E
278	262	-	-	E
279	263	-	-	-
280	264	-	-	-
281	265	-	-	E
282	266	-	-	-
283	267	-	-	-
284	268	-	-	-
285	269	-	-	-
286	270	-	-	-
287	271	-	-	-
288	272	-	-	-
289	273	-	-	C
290	274	-	-	-
291	275	C	C	C
292	276	-	-	C
293	277	-	-	-
294	278	C	C	C
295	279	-	-	C
296	280	-	-	C
297	281	-	-	-
298	282	-	-	-
299	283	-	-	-
300	284	-	-	-
301	285	-	-	-
302	286	-	-	-
303	287	-	-	-
304	288	-	-	-
305	289	-	-	-
306	290	-	-	-
307	291	-	-	-
308	292	-	-	-
309	293	-	-	-
310	294	-	-	C
311	295	-	-	-
312	296	-	-	-
313	297	-	-	C
314	298	-	-	-
315	299	-	-	C
316	300	-	-	C

317	301	-	-	-
318	302	-	-	-
319	303	-	-	-
320	304	-	-	C
321	305	-	-	C
322	306	-	-	-
323	307	-	N	C
324	308	-	-	C
325	309	-	-	C
326	310	-	-	C
327	311	-	-	C
328	312	-	-	C
329	313	-	-	-
330	314	-	-	-
331	315	-	-	-
332	316	-	-	-
333	317	-	-	-
334	318	-	-	-
335	319	-	-	-
336	320	-	-	-
337	321	-	-	-
338	322	-	-	-
339	323	-	N	-
340	324	-	-	-
341	325	-	-	-
342	326	-	-	-
343	327	N	N	-
344	328	-	-	-
345	329	-	-	-
346	330	-	-	-
347	331	-	-	-
348	332	-	-	-
349	333	-	-	-
350	334	-	-	-
351	335	-	-	-
352	336	-	-	-
353	337	-	-	-
354	338	-	-	-
355	339	-	-	-
356	340	-	-	-
357	341	-	-	-
358	342	-	-	-
359	343	-	-	-
360	344	-	-	-
361	345	-	-	-
362	346	-	-	-
363	347	-	-	-
364	348	-	-	-
365	349	-	-	-

366	350	-	-	-
367	351	-	-	-
368	352	-	-	-
369	353	-	-	-
370	354	-	-	-
371	355	-	-	-
372	356	-	-	-
373	357	-	-	-
374	358	-	-	-
375	359	-	-	-
376	360	-	-	-
377	361	-	-	-
378	362	-	-	-
379	363	-	-	-
380	364	-	-	-
381	365	-	-	-
382	366	-	-	-
383	367	-	-	-
384	368	-	-	-
385	369	-	-	-
386	370	-	-	-
387	371	-	-	-
388	372	-	-	-
389	373	-	-	-
390	374	-	-	-
391	375	-	-	-
392	376	-	-	-
393	377	-	-	-
394	378	-	-	-
395	379	-	-	-
396	380	-	-	-
397	381	-	-	-
398	382	-	-	-
399	383	-	-	-
400	384	-	-	-
401	385	-	-	-
402	386	-	-	-
403	387	-	-	-
404	388	-	-	-
405	389	-	-	-
406	390	-	-	-
407	391	-	-	-
408	392	-	-	-
409	393	-	-	-
410	394	-	-	-
411	395	-	-	-
412	396	-	-	-
413	397	-	-	-
414	398	-	-	-

415	399	-	-	-
416	400	-	-	-
417	401	-	-	-
418	402	-	-	-
419	403	-	-	-
420	404	-	-	-
421	405	-	-	-
422	406	-	-	-
423	407	-	-	-
424	408	-	-	-
425	409	-	-	-
426	410	-	-	-
427	411	-	-	-
428	412	-	-	-
429	413	-	-	-
430	414	-	-	-
431	415	-	-	-
432	416	-	-	-
433	417	-	-	-
434	418	-	-	-
435	419	-	-	-
436	420	-	-	-
437	421	-	-	-
438	422	-	-	-
439	423	-	-	-
440	424	-	-	-
441	425	-	-	-
442	426	-	-	-
443	427	-	-	-
444	428	-	-	-
445	429	-	-	-
446	430	-	-	-
447	431	-	-	-
448	432	-	-	-
449	433	-	-	-
450	434	-	-	-
451	435	-	-	-
452	436	-	-	-
453	437	-	-	-
454	438	-	-	-
455	439	-	-	-
456	440	-	-	-
457	441	-	-	-
458	442	-	-	-
459	443	-	-	-
460	444	-	-	-
461	445	-	-	-
462	446	-	-	-
463	447	-	-	-

464	448	-	-	-
465	449	-	-	-
466	450	-	-	-
467	451	-	-	-
468	452	-	-	-
469	453	-	-	-
470	454	-	-	-
471	455	-	-	-
472	456	-	-	-
473	457	-	-	-
474	458	-	-	-
475	459	-	-	-
476	460	-	-	-
477	461	-	-	-
478	462	-	-	-
479	463	-	-	-
480	464	-	-	-
481	465	-	-	-
482	466	-	-	-
483	467	-	-	-
484	468	-	-	-
485	469	-	-	-
486	470	-	-	-
487	471	-	-	-
488	472	-	-	-
489	473	-	-	-
490	474	-	-	-
491	475	-	-	-
492	476	-	-	-
493	477	-	-	-
494	478	-	-	-
495	479	-	-	-
496	480	-	-	-
497	481	-	-	-
498	482	-	-	-
499	483	-	-	-
500	484	-	-	-
501	485	-	-	-
502	486	-	-	-
503	487	-	-	-
504	488	-	-	-
505	489	-	-	-
506	490	-	-	-
507	491	-	-	-
508	492	-	-	-
509	493	-	-	-
510	494	-	-	-
511	495	-	-	-
512	496	-	-	-

513	497	-	-	-
514	498	-	-	-
515	499	-	-	-
516	500	-	-	-
517	501	-	-	-
518	502	-	-	-
519	503	-	-	-
520	504	-	-	-
521	505	-	-	-
522	506	-	-	-
523	507	-	-	-
524	508	-	-	-
525	509	-	-	-
526	510	-	-	-
527	511	-	-	-
528	512	-	-	-
529	513	-	-	-
530	514	-	-	-
531	515	-	-	-
532	516	-	-	-
533	517	-	-	-
534	518	-	-	-
535	519	-	-	-
536	520	-	-	-
537	521	-	-	-
538	522	-	-	-
539	523	-	-	-
540	524	-	-	-
541	525	-	-	-
542	526	-	-	-
543	527	-	-	-
544	528	-	-	-
545	529	-	-	-
546	530	-	-	-
547	531	-	-	-
548	532	-	-	-
549	533	-	-	-
550	534	-	-	-
551	535	-	-	-
552	536	-	-	-
553	537	-	-	-
554	538	-	-	-
555	539	-	-	-
556	540	-	-	-
557	541	-	-	-
558	542	-	-	-
559	543	-	-	-
560	544	-	-	-
561	545	-	-	-

562	546	-	-	-
563	547	-	-	-
564	548	-	-	-
565	549	-	-	-
566	550	-	-	-