Published in final edited form as:

Science. 2019 May 03; 364(6439): 480–484. doi:10.1126/science.aav5095.

Structural topology defines protective CD8⁺ T cell epitopes in the HIV proteome

Gaurav D. Gaiha^{1,2,*}, Elizabeth J. Rossin^{3,4,*}, Jonathan Urbach¹, Christian Landeros¹, David R. Collins^{1,5}, Chioma Nwonu¹, Itai Muzhingi¹, Melis N. Anahtar^{1,6}, Olivia M. Waring^{1,7}, Alicja Piechocka-Trocha^{1,5}, Michael Waring^{1,5}, Daniel P. Worrall¹, Musie S. Ghebremichael¹, Ruchi M. Newman¹, Karen A. Power¹, Todd M. Allen¹, James Chodosh⁴, Bruce D. Walker^{1,3,5,7,†}

¹Ragon Institute of MGH, MIT and Harvard, Cambridge, MA 02139, USA.

²Gastrointestinal Unit, Massachusetts General Hospital, Boston, MA 02114, USA.

³The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

⁴Department of Ophthalmology, Massachusetts Eye and Ear, Boston, MA 02114, USA.

⁵Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA.

⁶Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA.

⁷Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

Abstract

Mutationally constrained epitopes of variable pathogens represent promising targets for vaccine design but are not reliably identified by sequence conservation. In this study, we employed structure-based network analysis, which applies network theory to HIV protein structure data to quantitate the topological importance of individual amino acid residues. Mutation of residues at important network positions disproportionately impaired viral replication and occurred with high frequency in epitopes presented by protective human leukocyte antigen (*HLA*) class I alleles.

† Corresponding author. bwalker@mgh.harvard.edu.

*These authors contributed equally to this work.

Author contributions: G.D.G., E.J.R., and B.D.W. conceived of and designed the study. G.D.G. and E.J.R. conceptualized the structure-based network analysis methodology. E.J.R. designed, wrote, and validated the software. G.D.G., C.L., C.N., D.R.C., I.M., A.P.-T., M.W., R.M.N., and K.A.P. performed the experiments. G.D.G., E.J.R., J.U., D.R.C., M.N.A., O.M.W., M.S.G., R.M.N., K.A.P., and T.M.A. performed computational analyses. D.P.W. managed patient recruitment and sample collection. J.U. generated protein network diagrams. B.D.W. and J.C. supervised the study. G.D.G., E.J.R., and B.D.W. wrote the manuscript. All authors read and approved the manuscript.

Competing interests: G.D.G., E.J.R., and B.D.W. have filed a provisional patent application (62/817,094).

Data and materials availability: All mutational tolerance data are available in the original publications (see table S1). Viral sequence data are available at the SRA database (accession number PRJNA524922). All other data are available in the manuscript or supplementary materials. Code is archived at Zenodo (31). The following reagents were obtained through the AIDS Research and Reference Program, Division of AIDS, NIAID, NIH: pNL4-3 from M. Martin and TZM-bl cells from J. Kappes and X. Wu.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6439/480/suppl/DC1

Materials and Methods

Figs. S1 to S9

Tables S1 to S4

References (32–64)

Moreover, CD8⁺ T cell targeting of highly networked epitopes distinguished individuals who naturally control HIV, even in the absence of protective *HLA* alleles. This approach thereby provides a mechanistic basis for immune control and a means to identify CD8⁺ T cell epitopes of topological importance for rational immunogen design, including a T cell-based HIV vaccine.

The development of an effective HIV vaccine is a critical global health priority (1). An important component of this effort is defining immune responses of individuals who exhibit natural viral control (2). Genome-wide association studies (GWAS) have identified strong associations of HIV control with certain human leukocyte antigen (*HLA*) class I alleles (e.g., *B*57* and *B*27*) and specific amino acids lining the HLA-peptide binding pocket (3, 4), suggesting a key role for cytotoxic T lymphocyte (CTL) epitope specificity. However, the extent to which targeting specific epitopes influences viral control, and the distinguishing features of protective epitopes, remain poorly understood.

Sequence-conserved epitopes have been considered optimal targets of efficacious CTL responses (5, 6), but targeting conserved epitopes is not distinctively associated with immune control (7). Only a subset of sequence-conserved residues exact a substantial change in viral fitness when mutated (8, 9). Higher-order sequence analysis or quantitative fitness landscapes can predict regions of vulnerability, as they capture effects of epistasis between protein residues more accurately than viral sequence entropy (10–12). Although it has been suggested that epistatic effects may originate from structural constraints, systematic means to directly evaluate viral protein structure and quantitate mutational constraints have not been defined.

To address this, we developed structure-based network analysis, which uses protein structure data and network theory to quantify the topological importance of each amino acid residue to tertiary and quaternary protein structure. We used atomic-level coordinate data from the Protein Data Bank (PDB; www.rcsb.org) to build networks of amino acids (nodes) and noncovalent interactions (edges), including van der Waals interactions, hydrogen bonds, salt bridges, disulfide bonds, π - π interactions, π -cation interactions, metal-coordinated bonds, and local hydrophobic packing, each of which captures distinct aspects of residue-residue interactions. Using this network-based representation, we calculated an array of network centrality metrics (measures of relative importance in a given network topology) (13). This provided a quantitative measure of topological importance for each amino acid residue (i.e., a network score) through an assessment of local connectivity, involvement as bridges between higher-order protein elements, and proximity to protein ligands (fig. S1 and materials and methods).

We validated this approach on 13 proteins with published functional outcomes from high-throughput mutagenesis experiments (table S1), revealing significant inverse correlations between network scores and experimentally derived mutational tolerance values ($P = 6.9 \times 10^{-4}$ to 2.6×10^{-66}) (Fig. 1A and fig. S2). Although there was a range of correlation coefficients (Spearman's $\rho = -0.46$ to -0.72), some of the strongest correlations were between network scores and experimental data linked to an essential process (e.g., β -lactamase and ampicillin resistance), suggesting a robust relationship between residue topology and functional importance. The analysis also identified residues of low mutational

tolerance (bottom 10%) more accurately than did sequence conservation or relative solvent accessibility (RSA) (Fig. 1B and fig. S3).

We therefore applied this approach to assess mutational constraints within the HIV proteome. Residue network scores for monomeric (Fig. 1C) and higher-order Gag p24 conformations were binned into quintiles and compared with sequence entropy values from 5430 clade B isolates. This revealed a strong inverse relationship between network measures of topological importance and mutational frequency (Fig. 1D). Extending the analysis to 11 additional HIV proteins illustrated a similar inverse relationship (Fig. 1E), indicating that this finding was broadly applicable across the HIV proteome.

To experimentally evaluate the relationship between network score and mutational tolerance, we engineered point mutations into a set of conserved residues with high (>1) or low (<1) network scores, and a set of matched non-conserved, low network score residues, in Gag p24, reverse transcriptase, integrase, and gp120 (table S2). Mutation of highly networked residues led to substantial impairment of HIV infectivity at 2 days and viral replication at 7 days (Fig. 1, F to H). In contrast, mutation of residues with low network scores, regardless of conservation, had little impact. Moreover, there was a strong inverse correlation between viral infectivity and network score, but not sequence entropy (Fig. 1, I and J).

Comparative assessment of individual HIV proteins by a standardized network metric (second-order degree centrality) allowed us to rank the proteins on the basis of median residue connectivity. This revealed that Gag p24 was statistically the most highly networked protein, particularly in comparison to envelope and accessory proteins (Fig. 2A). The higher second-order degree centrality of Gag p24 is a result of the extensive multimerization necessary for capsid formation (fig. S4A), which potentially elucidates both the mutational fragility of Gag p24 (14) and observations linking Gag-specific CTL breadth with lower viral loads (15).

We next evaluated network scores of a representative risk allele *B*35Px* epitope and protective allele *B*57* epitope (4, 16, 17). The *B*35-DL9* epitope (gp120 78–86) contained poorly networked residues (Fig. 2B), whereas the *B*57-KF11* epitope (Gag p24 30–40) contained highly networked residues that bridge the N- and C-terminal domains of Gag p24 (Fig. 2, C and D), consistent with its limited capacity to escape CTL pressure (18). Furthermore, highly networked residues in KF11 occupied immunologically important HLA anchor and T cell receptor (TCR) contact sites, suggesting a mechanism for durable epitope presentation and CTL recognition.

We next computed epitope network scores for all optimally defined CTL epitopes with high-quality structural data (~89.2%) (19). We accomplished this by summing average residue network scores involved in HLA binding, TCR recognition, and peptide processing as three discrete quantities, to evenly capture the described mechanisms of viral epitope escape (table S3 and materials and methods) (20–22). This revealed that median epitope network scores of protective *HLA* alleles, as defined by GWAS (4), were statistically higher than those of neutral or risk alleles (Fig. 2, E and F) and became further differentiated when comparisons were limited to immunodominant epitopes (Fig. 2G) (16). Median epitope network scores

were also positively correlated with GWAS-defined odds ratios (Fig. 2H). Closer examination of *HLA-B*57* revealed that a high percentage of its epitopes were highly networked (within the top decile, ~27.2%) (fig. S5), consistent with its enrichment in HIV controllers across diverse cohorts (3, 4, 23). The amino acid with the highest median network score was tryptophan, which is a common C-terminal *HLA* anchor for *B*57*-restricted epitopes but not for other *HLA* alleles (Fig. 2I). Indeed, tryptophan anchors in *B*57* epitopes were among the highest scoring throughout the HIV proteome.

In addition to distinguishing protective from risk *HLA* alleles, the analysis identified several topologically important epitopes restricted by neutral *HLA* alleles (table S3). To determine whether HIV control can be mediated by targeting highly networked epitopes irrespective of protective *HLA* allele expression, we evaluated the proliferative CD8⁺ T cell responses of 114 untreated HIV-positive individuals with elite or viremic control (viral load < 2000 copies/ml), intermediate viral loads (2000 to 10,000 copies/ml), or viral progression (viral load > 10,000 copies/ml) (table S4). We focused on CTL proliferation, given its strong association with CTL functionality and immune control (24, 25).

Evaluation of a representative controller and progressor revealed a substantial difference in network scores of immunodominant epitopes (Fig. 3, A to C). The KL9 epitope (gp120 121–129) targeted by the controller and restricted by the neutral *HLA-A*02* allele contained highly networked residues from the V1V2 stem of gp120 (Fig. 3D) (26). In contrast, the *B*07*-RI10 epitope (gp120 298–307) targeted by the progressor (Fig. 3B) contained low-scoring residues from the V3 loop in both open and closed conformations of trimeric gp120 (Fig. 3, C and D) (27, 28). Overall, controllers preferentially targeted epitopes with high network scores with their strongest responses, whereas individuals with intermediate or high viral loads had weak or absent responses against these epitopes (Fig. 3E and fig. S6). Statistical analysis confirmed that epitope network scores significantly differentiated controllers from progressors (Fig. 3F) and were superior to a complementary sequence conservation metric (fig. S8). Although these individuals could be differentiated by CTL proliferation (Fig. 3G), they became further distinguished when the magnitude of CTL proliferation was incorporated for each epitope-specific response (Fig. 3H), implicating the importance of both CTL function and specificity in HIV control. We observed a strong inverse correlation between summed epitope network scores and viral load (Spearman's $\rho = -0.63$, $P < 0.0001$) (fig. S8).

A subanalysis comparing controllers with protective or nonprotective *HLA* alleles revealed no significant difference in epitope network scores when analyzed alone (Fig. 3I) or when CTL proliferation was incorporated (Fig. 3J). To control for protective *HLA* allele expression, we also evaluated *HLA-B*57*⁺ controllers and progressors and observed a significant difference in the network scores of targeted epitopes scaled by CTL proliferation (Fig. 3K). A representative example involves an *HLA-B*57*⁺ controller targeting the highly networked IW9 epitope (Gag p24 15–23) (29) and an *HLA-B*57*⁺ progressor targeting the poorly networked Vif IF9 epitope (31–39) (Fig. 3M). To control for T cell functionality, we compared a subset of controllers and progressors with similar levels of CTL proliferation (Fig. 3N) and found a significant difference in the sum of network scores of targeted epitopes (Fig. 3O). Collectively, these subanalyses demonstrated that CTL targeting of

highly networked epitopes was a key component of immune control across diverse *HLA* alleles.

To evaluate the evolutionary constraints of epitopes with high and low network scores, we performed plasma viral sequence analysis of epitopes targeted by nine controllers and 15 progressors with similar proliferative CTL responses. A representative example of two epitopes from Nef (Fig. 4, A and D) revealed no sequence variation in the high-scoring *B*53:YF9* epitope (Fig. 4, B and C) but numerous mutations in the low-scoring *B*08:FL8* epitope (Fig. 4, E and F), which abrogated subsequent CTL recognition (fig. S9). Cumulative assessment of epitope sequence data from controllers and progressors revealed statistically significant differences in mutation frequency (Fig. 4H), particularly at HLA anchor and TCR contact sites (Fig. 4I). Notably, only three of the nine controllers targeted epitopes restricted by protective *HLA* alleles.

In this Report, we applied structure-based network analysis to define the topological importance of residues and CTL epitopes across the HIV proteome. This analysis suggests that functional CTL targeting of epitopes that contain topologically important viral residues at HLA anchor, TCR contact, and flanking epitope sites is a broad mechanism of immune control. Moreover, identification of high network score epitopes presented by major *HLA* supertypes (30) provides the basis for a rational T cell-based HIV vaccine with global coverage. These data suggest that prophylactic and therapeutic vaccination should aim, in part, to induce functional CTL responses against topologically important epitopes, which constitute a subset of conserved epitopes that may not be commonly targeted during natural infection. With the ever-expanding supply of high-resolution crystal structures, this approach could be broadly applied to a diverse array of pathogens. This methodology therefore represents a promising tool by which rational structural vaccinology can be used to enhance the design of T cell-based vaccines.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank M. Carrington, V. Naranbhai, A. Chakraborty, and Z. Brumme for their advice and comments on the manuscript.

Funding: Support was provided by the Howard Hughes Medical Institute (B.D.W.); CHAVI-ID (B.D.W.); the Ragon Institute of MGH, MIT and Harvard (B.D.W.); NIH HIVRAD grant P01AI104715 (T.M.A.); Mark and Lisa Schwartz; the Harvard University Center for AIDS Research (P30 AI060354 to B.D.W.), which is supported by the following institutes and centers co-funded by and participating with the U.S. National Institutes of Health: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, and OAR; a CFAR Development Award (G.D.G.); and a Harvard Department of Ophthalmology Gragoudas-Folkman Award (E.J.R.).

REFERENCES AND NOTES

1. Walker BD, in The Vaccine Book, Bloom BR, Lambert P-H, Eds. (Academic Press, ed. 2, 2016), pp. 401–422.
2. Deeks SG, Walker BD, Immunity 27, 406–416 (2007). [PubMed: 17892849]
3. Fellay J et al., Science 317, 944–947 (2007). [PubMed: 17641165]

4. International HIV Controllers Study, *Science* 330, 1551–1557 (2010). [PubMed: 21051598]
5. Rolland M, Nickle DC, Mullins JI, *PLOS Pathog.* 3, e157 (2007). [PubMed: 18052528]
6. Létourneau S et al., *PLOS ONE* 2, e984 (2007). [PubMed: 17912361]
7. Migueles SA et al., *EBioMedicine* 2, 46–58 (2014). [PubMed: 26137533]
8. Rolland M et al., *J. Virol.* 87, 5461–5467 (2013). [PubMed: 23468488]
9. Manochewa S, Swain JV, Lanxon-Cookson E, Rolland M, Mullins JI, *PLOS ONE* 8, e66065 (2013). [PubMed: 23785468]
10. Dahirel V et al., *Proc. Natl. Acad. Sci. U.S.A.* 108, 11530–11535 (2011). [PubMed: 21690407]
11. Ferguson L et al., *Immunity* 38, 606–617 (2013). [PubMed: 23521886]
12. Barton JP et al., *Nat. Commun.* 7, 11660 (2016). [PubMed: 27212475]
13. Borgatti SP, *Soc. Networks* 27, 55–71 (2005).
14. Rihn SJ et al., *PLOS Pathog.* 9, e1003461 (2013). [PubMed: 23818857]
15. Kiepiela P et al., *Nat. Med.* 13, 46–53 (2007). [PubMed: 17173051]
16. Streeck H et al., *J. Virol.* 83, 7641–7648 (2009). [PubMed: 19458000]
17. Gao X et al., *N. Engl. J. Med.* 344, 1668–1675 (2001). [PubMed: 11386265]
18. Gorin M et al., *PLOS Pathog.* 13, e1006541 (2017). [PubMed: 28787455]
19. Llano, Williams A, Olvera A, Silva-Arrieta S, Brander C, in *HIV Molecular Immunology 2013*, Yusim K et al., Eds. (Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR 13–27758, 2013), pp. 3–25.
20. Yang OO et al., *J. Exp. Med.* 197, 1365–1375 (2003). [PubMed: 12743169]
21. Draenert R et al., *J. Exp. Med.* 199, 905–915 (2004). [PubMed: 15067030]
22. Bronke et al., *AIDS* 27, 899–905 (2013). [PubMed: 23276808]
23. Kiepiela P et al., *Nature* 432, 769–775 (2004). [PubMed: 15592417]
24. Ndhlovu ZM et al., *Blood* 121, 801–811 (2013). [PubMed: 23233659]
25. Migueles SA et al., *Immunity* 29, 1009–1021 (2008). [PubMed: 19062316]
26. O’Connell RJ, Kim JH, Excler J-L, *Expert Rev. Vaccines* 13, 1489–1500 (2014). [PubMed: 25163695]
27. Ward B, Wilson IA, *Trends Biochem. Sci.* 40, 101–107 (2015). [PubMed: 25600289]
28. Wang H et al., *Proc. Natl. Acad. Sci. U.S.A.* 113, E7151–E7158 (2016). [PubMed: 27799557]
29. Pornillos O, Ganser-Pornillos BK, Yeager M, *Nature* 469, 424–427 (2011). [PubMed: 21248851]
30. Sette J Sidney, *Immunogenetics* 50, 201–212 (1999). [PubMed: 10602880]
31. The Walker Lab at The Ragon Institute of Harvard, MIT, and MGH, O. Waring, WalkerLabRagon/NetworkAnalysis: Network Analysis Pipeline, Version 1.0, Zenodo (2019); 10.5281/zenodo.2597484.

Fig. 1. Structure-based network analysis identifies amino acid residues with low mutational tolerance.

(A) Scatter plot of average mutational tolerance and network score for TEM-1 β -lactamase residues (red, active site residues Ser⁷⁰, Lys⁷³, Glu¹⁶⁶, and Asn¹⁷⁰). (B) Comparative receiver operator curves (ROC) and area under the curve (AUC) characteristics for network score, RSA, and sequence entropy to identify the bottom 10% of residues of low mutational tolerance in TEM-1 β -lactamase. (C) Structure-based network schematic for Gag p24 monomer (PDB ID: 3J34, chain C), including amino acid residues (nodes) and noncovalent interactions (edges). Edge width indicates interaction strength and node size indicates relative network score. (D and E) Comparison of Gag p24 and HIV proteome network scores (binned by quintile: low, 2nd, 3rd, 4th, and high; top 5% in gray) with viral sequence entropy. (F) Comparison of viral infectivity of TZM-bl cells and (G and H) viral spreading within GXR cells after mutation of conserved, highly networked residues (blue); conserved, poorly networked residues (red); and nonconserved, poorly networked residues (green). Statistical analyses by one-way analysis of variance and Wilcoxon matched pairs test. (I and J) Scatter plot of sequence conservation and network scores with mutant virus infectivity. Correlations were calculated by Spearman's rank correlation coefficient. Statistical comparisons were made using Mann-Whitney *U* test. For comparisons of more than two groups, Kruskal-Wallis test with Dunn's post hoc analyses was used. Calculated *P* values are as follows: NS, not significant; **P* < 0.05; ***P* < 0.01; ****P* < 0.001; *****P* < 0.0001.

Fig. 2. Network score distinguishes HIV proteins and CTL epitopes associated with protective, neutral, and risk-associated HLA alleles.

(A) Ranked median second-order degree centrality values (red dots) for HIV proteins (median, interquartile range). (B and C) Network scores for risk and protective allele epitopes *B*35Px*-DL9 and *B*57*-KF11 (HLA anchor residues, red; TCR contact residues, blue). Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (D) Network-based depiction of *B*35Px*-DL9 in open gp120 (left, red) (PDB ID: 3J70, chain D) and *B*57*-KF11 in monomeric Gag p24 (right, blue) (PDB ID: 3J34, chain C). (E) Ranked median epitope network scores for individual HLA alleles (median, interquartile range). GWAS-defined protective and risk HLA alleles indicated in blue and red, respectively. (F) Comparison of epitope network scores presented by protective, neutral, and risk HLA alleles. (G) Comparison of epitope network scores of immunodominant epitopes presented by HLA alleles associated with protection (blue; *B*5701*-TW10, *B*5201*-RI8, *B*2705*-KK10, *B*1402*-DA9) and risk (red; *B*0801*-FL8, *B*3501*-DL9, *B*0702*-RV9, *Cw*07*-RY11). (H) Scatter plot of GWAS-defined protective HLA allele odds ratios (OR) to median epitope network score. (I) Ranked network scores of each amino acid type across the HIV proteome (median, interquartile range). HLA-*B*57* anchor residues denoted in blue. Statistical comparisons were made using Mann-Whitney *U* test. For comparisons of more than two groups, Kruskal-Wallis test with Dunn's post hoc analyses was used. Correlations were calculated by Spearman's rank correlation coefficient. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

Fig. 3. Targeting of topologically important epitopes distinguishes HIV controllers from progressors irrespective of HLA allele.

(A and B) Proliferative CTL responses in a single representative controller (top) and progressor (bottom) after 6-day incubation of carboxyfluorescein succinimidyl ester (CFSE)-loaded peripheral blood mononuclear cells with optimal epitopes matched to the person's HLA haplotype. VL, viral load. (C) Residue network scores of epitopes targeted in (B) (red, HLA anchor; blue, TCR contact). (D) Network-based depiction of A*02-KL9 and B*07-RI10 in open (top; PDB ID: 3J70, chain D) and closed (bottom; PDB ID: 5T3X) gp120 trimeric conformations. A single subunit of trimeric, open gp120 is presented for ease of epitope visualization. (E) Proliferative responses of controllers (blue), intermediates (green), and progressors (red). The x axis depicts all CTL epitopes ranked by epitope network score from lowest to highest. (F to H) Comparison of controllers (C; $n = 46$ individuals), intermediates (I; $n = 25$), and progressors (P; $n = 43$) by summed epitope network scores, summed proliferative responses, and summed epitope network scores scaled by proliferative CTL response. (I and J) Comparison of controllers with nonprotective alleles (NPC; $n = 18$) and protective alleles (PC; $n = 28$) by summed epitope network scores (I) and summed epitope network scores scaled by magnitude of proliferative CTL response (J).

(J). **(K)** Comparison of B^*57^+ controllers (C; $n = 17$) and B^*57^+ progressors (P; $n = 7$) by summed epitope network scores scaled by magnitude of CTL proliferation. **(L)** CFSE dilution of immunodominant CTL responses from a single representative B^*57^+ controller and B^*57^+ progressor. **(M)** Residue network scores of CTL epitopes targeted in (L). **(N and O)** Comparison of controllers (C; $n = 14$) and progressors (P; $n = 14$) with similar magnitude of summed CTL proliferation (upper) by CTL proliferation (N) and epitope network scores (O). Statistical comparisons were made using Mann-Whitney U test. For comparisons of more than two groups, Kruskal-Wallis test with Dunn's post hoc analyses was used. $*P < 0.05$; $**P < 0.01$; $***P < 0.001$; $****P < 0.0001$.

Fig. 4. Topologically important CTL epitopes targeted by HIV controllers are infrequently mutated in vivo.

(A and D) CFSE dilution of immunodominant CTL responses from a single representative controller and progressor. (B and E) Network scores of the *B*53*-YF9 and *B*08*-FL8 epitopes (red, HLA anchor; blue, TCR contact). (C and F) WebLogo of *B*53*-YF9 and *B*08*-FL8 sequence data (red, HLA anchor; blue, TCR contact; green, flanking). (G) Network representation of the *B*08*-FL8 and *B*53*-YF9 epitopes within the Nef dimer (PDB ID: 2XI1). (H) Comparison of average number of mutations within epitopes targeted by controllers ($n = 9$) and progressors ($n = 15$) for each individual patient. (I) Comparison of the percent frequency of mutations at HLA anchor (blue), TCR contact (red), and flanking residues (black) between controller-targeted (open bars) and progressor-targeted epitopes (filled bars). Statistical comparisons were made using Mann-Whitney U test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.