



















## ARTICLE

# A cell-free antigen processing system informs HIV-1 epitope selection and vaccine design

Srona Sengupta<sup>1,2,8</sup> , Josephine Zhang<sup>1\*</sup> , Madison C. Reed<sup>1\*</sup> , Jeanna Yu<sup>1\*</sup> , Aeryon Kim<sup>3</sup> , Tatiana N. Boronina<sup>4</sup> , Nathan L. Board<sup>1</sup> , James O. Wrabl<sup>5</sup> , Kevin Shenderov<sup>1</sup> , Robin A. Welsh<sup>2</sup> , Weiming Yang<sup>2</sup> , Andrew E. Timmons<sup>1</sup> , Rebecca Hoh<sup>6</sup> , Robert N. Cole<sup>4</sup> , Steven G. Deeks<sup>6</sup> , Janet D. Siliciano<sup>1</sup> , Robert F. Siliciano<sup>1,7</sup> , and Scheherazade Sadegh-Nasseri<sup>2</sup> 

**Distinct CD4<sup>+</sup> T cell epitopes have been associated with spontaneous control of HIV-1 replication, but analysis of antigen-dependent factors that influence epitope selection is lacking. To examine these factors, we used a cell-free antigen processing system that incorporates soluble HLA-DR (DR1), HLA-DM (DM), cathepsins, and full-length protein antigens for epitope identification by LC-MS/MS. HIV-1 Gag, Pol, Env, Vif, Tat, Rev, and Nef were examined using this system. We identified 35 novel epitopes, including glycopeptides. Epitopes from smaller HIV-1 proteins mapped to regions of low protein stability and higher solvent accessibility. HIV-1 antigens associated with limited CD4<sup>+</sup> T cell responses were processed efficiently, while some protective epitopes were inefficiently processed. 55% of epitopes obtained from cell-free processing induced memory CD4<sup>+</sup> T cell responses in HIV-1<sup>+</sup> donors, including eight of 19 novel epitopes tested. Thus, an in vitro processing system utilizing the components of Class II processing reveals factors influencing epitope selection of HIV-1 and represents an approach to understanding epitope selection from non-HIV-1 antigens.**

## Introduction

CD4<sup>+</sup> T cells connect the humoral- and cell-mediated arms of the immune system, both of which are vital for vaccine responses against chronic viral infections. Elegant analyses of CD4<sup>+</sup> T cell responses to HIV-1 by Walker and colleagues have provided insights into protective HLA-DR alleles and viral epitopes associated with control of viral replication. The breadth and magnitude of Gag-specific CD4<sup>+</sup> T cell responses are associated with anti-Env neutralizing antibodies (Ranasinghe et al., 2015) and inversely correlated with viral load (Laher et al., 2017; Ranasinghe et al., 2012; Ranasinghe et al., 2013). Three CD4<sup>+</sup> T cell epitopes in Gag are associated with spontaneous viral control (Ranasinghe et al., 2012, 2013). However, it remains unclear why few individuals develop these protective responses and whether this is influenced by antigen-processing mechanisms.

The antigen-processing pathway for major histocompatibility complex Class II (MHC-II)-restricted CD4<sup>+</sup> T cell epitopes begins with the endocytosis of exogenous antigens or autophagy of intracellular contents (Unanue et al., 2016). Full-length protein antigens bind to MHC-II molecules in the late endosomal MHC-II compartment (MIIC) of professional APCs, with subsequent

cleavage/trimming of the exposed protein around the bound region (Kim et al., 2014). Epitope selection by MHC-II is facilitated by chaperones HLA-DM (DM) and HLA-DO. DM recognizes structurally flexible conformations of peptide:MHC-II (pMHC-II) complexes (Chou and Sadegh-Nasseri, 2000) due to partially filled or unfilled P1 pockets in the peptide binding groove (Anders et al., 2011; Chou and Sadegh-Nasseri, 2000) and induces dissociation of poorly bound peptides such as the class II-associated invariant chain (CLIP). The resulting open MHC-II groove is peptide-receptive and can rapidly scan for the best fitting sequences (Chou and Sadegh-Nasseri, 2000; Natarajan et al., 1999; Rabinowitz et al., 1998), generating tightly formed pMHC-II complexes that are no longer recognized by DM (Narayan et al., 2007, 2009) and that would likely have a longer half-life on the surface of APCs (Nelson et al., 1994). DO works cooperatively with DM to promote its function in B cells, where it is primarily expressed (Poluektov et al., 2013; Welsh et al., 2019, 2020).

CD4<sup>+</sup> T cell epitopes are often defined using overlapping synthetic peptides. While this approach provides a broad survey of the T cell epitopes within proteins of interest, it does not

<sup>1</sup>Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; <sup>2</sup>Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA; <sup>3</sup>Department of Inflammation and Oncology and Genome Analysis Unit, Amgen Research, Amgen Inc., South San Francisco, CA, USA; <sup>4</sup>Department of Biological Chemistry, Mass Spectrometry and Proteomics Facility, Johns Hopkins University School of Medicine, Baltimore, MD, USA; <sup>5</sup>Department of Biology, Johns Hopkins University, Baltimore, MD, USA; <sup>6</sup>Department of Medicine, University of California, San Francisco, San Francisco, CA, USA; <sup>7</sup>Howard Hughes Medical Institute, Baltimore, MD, USA; <sup>8</sup>The Graduate Program in Immunology and Medical Scientist Training Program, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

\*J. Zhang, M.C. Reed, and J. Yu contributed equally to this paper. Correspondence to Scheherazade Sadegh-Nasseri: [ssadegh@jhmi.edu](mailto:ssadegh@jhmi.edu); Robert F. Siliciano: [rsiliciano@jhmi.edu](mailto:rsiliciano@jhmi.edu).

reveal the parameters that determine epitope selection. This is because the antigen-processing steps needed to generate pathogen-derived pMHC-II complexes are bypassed. CD4<sup>+</sup> T cell responses are the final output in a long series of steps following infection and are influenced by antigen-dependent and independent factors. Antigen-dependent factors include structural features of the native protein, the molecular context of the epitope within the protein (Kim and Sadegh-Nasseri, 2015; Kim et al., 2017; Mirano-Bascos et al., 2008), the affinity of the epitope for the relevant MHC molecule (Yewdell and Bennink, 1999), resistance to DM-mediated editing of the pMHC-II complex (Kim and Sadegh-Nasseri, 2015) leading to the selection of stable pMHC-II (Lazarski et al., 2005), and TCR affinity for pMHC-II (Malherbe et al., 2004). Antigen-independent factors include the composition of the naive TCR repertoire (Jenkins and Moon, 2012; Kim et al., 2005; Moon et al., 2007), genetic polymorphisms (i.e., in T cell signaling, T cell-APC interactions, antigen processing genes, and host factors regulating viral replication), and prior exposure to crossreactive pathogens that may influence HIV-1-specific memory T cell frequencies (Campion et al., 2014; Su et al., 2013).

We have previously developed a reductionist cell-free antigen processing system that mimics the MIIC and yields immunodominant epitopes from full-length proteins that induce memory CD4<sup>+</sup> T cell responses. The system is composed of soluble MHC-II (HLA-DR1\*01:01, or DR1), DM, and three cathepsins in an acidic and reducing environment (Hartman et al., 2010; Kim et al., 2014), and as such, uniquely isolates structural and antigenic factors involved in epitope selection. Protein antigens are denatured in this environment, captured by DR1, further selected with the help of DM, and trimmed by cathepsins. DR1-bound peptides are then eluted and sequenced by liquid chromatography tandem mass spectrometry (LC-MS/MS). This system successfully identified the dominant epitopes of several pathogens and autoantigens in mice and humans (Hartman et al., 2010; Kim et al., 2014, 2017). We reasoned that such a system may reveal patterns of epitope selection for HIV-1, a pathogen for which most of our understanding of this process has derived from CD4<sup>+</sup> T cell responses to overlapping peptides (Lindqvist et al., 2012; Ranasinghe et al., 2012; Soghoian et al., 2012; Laher et al., 2017). Analysis of antigen-dependent factors leading to epitope selection across the HIV-1 proteome, including epitopes associated with viral control, is lacking. Thus, we used our cell-free processing system to understand how DM resistance and antigen structure influence epitope presentation from HIV-1 protein antigens. We hypothesized that our system could provide a broader landscape of potential HIV-1 epitopes and may identify novel epitopes not previously documented in studies of CD4<sup>+</sup> T cell responses to HIV-1. Vaccine-induced responses to these novel epitopes might contribute to vaccine efficacy even if the relevant epitopes are not normally selected for in infected individuals.

## Results

### Cell-free processing of HIV-1 proteome identifies MHC-II epitopes

HIV-1 proteins, as well as individual protein subunits derived from HIV-1 polyproteins, were incubated in a low pH (5.0–5.2)

reducing environment with DR1 with or without DM and then digested with cathepsins B, H, and S, followed by MS analysis (Fig. 1, A and B). This protocol mimics the natural conditions and sequence of events in MHC-II antigen processing (Hartman et al., 2010; Kim et al., 2014; Sadegh-Nasseri and Kim, 2015). We specifically allowed for antigen binding to DR1 first before adding cathepsins as this model is facilitated by the open-ended groove of MHC-II and supported by various studies (Castellino et al., 1998; Mimura et al., 2007; Nelson et al., 1997), including those showing that full-length reduced versions of antigens can bind to MHC-II molecules (Kim et al., 2014; Runnels et al., 1997; Sette et al., 1989) and that epitopes from viral proteins are susceptible to cathepsin digestion if not captured by DR1 (Kim et al., 2014). We performed cell-free processing experiments with and without DM, as resistance of a pMHC-II complex to DM-mediated dissociation predicts immunodominance for pathogen-derived proteins better than affinity of the peptide for the MHC or the intrinsic half-life of the complex (Hartman et al., 2010; Yin et al., 2012; Fig. 1, A and B). Epitopes identified in the presence of DM were considered DM-resistant and likely immunogenic (Fig. S1, A and B).

We subjected nearly the entire HIV-1 proteome to cell-free processing (Fig. 1 C and Table S1). Where possible, we performed processing of HIV-1 polyproteins and their individual subunits as various forms of an HIV-1 protein may be present in an infected CD4<sup>+</sup> T cell whose contents are captured by an APC (Addison et al., 2022). The HIV-1 Gag polyprotein (Pr55<sup>Gag</sup>) is cleaved by the viral protease to liberate the structural proteins matrix (MA/p17), capsid (CA/p24), and nucleocapsid (NC/p7) as well as the unstructured proteins spacer peptide 1 (SP1/p2), spacer peptide 2 (SP2/p1), and p6 (Fig. 1 C; Freed, 2015). We utilized the cell-free processing system on multiple Gag forms: p17, p24, and p24-p2-p7 precursor proteins (Table S1). We also examined the individual HIV-1 enzymes protease (PR), reverse transcriptase (RT), and integrase (INT), which are encoded by the *pol* gene. Despite their lower abundance relative to Gag in infected cells, their critical role in the viral life cycle and high conservation provide a strong impetus to identify immunogenic MHC-II Pol epitopes (Fig. 1 C). Additionally, we subjected monomeric gp120 (JR-FL strain) and a trimeric form of gp140, the extracellular portion of the Env protein (BG505 SOSIP.664; Sanders et al., 2013), to cell-free processing (Fig. 1 C and Table S1). Finally, we analyzed cell-free processing of all accessory proteins available commercially (Vif, Tat, Rev, and Nef; Fig. 1 C and Table S1).

Across the HIV-1 proteome, we observed clusters of nested epitopes containing overlapping sequences with typical CD4<sup>+</sup> T cell epitope length variation (Fig. 1 C). From 80 identified clusters that could bind to DR1, 55 were DM-resistant and 35 had not previously been reported (Fig. 1 D). Notably, all HIV-1 proteins subjected to in vitro processing—constituting the majority of the viral proteome—generated epitopes that could bind to a single MHC-II allele, DR1 (Fig. 1 C).

### Cell-free processing reveals hot spots of HIV-1 epitopes

Epitope “hot spots” were apparent in all proteins tested (Fig. 1 C). Cell-free processing of the myristoylated matrix protein

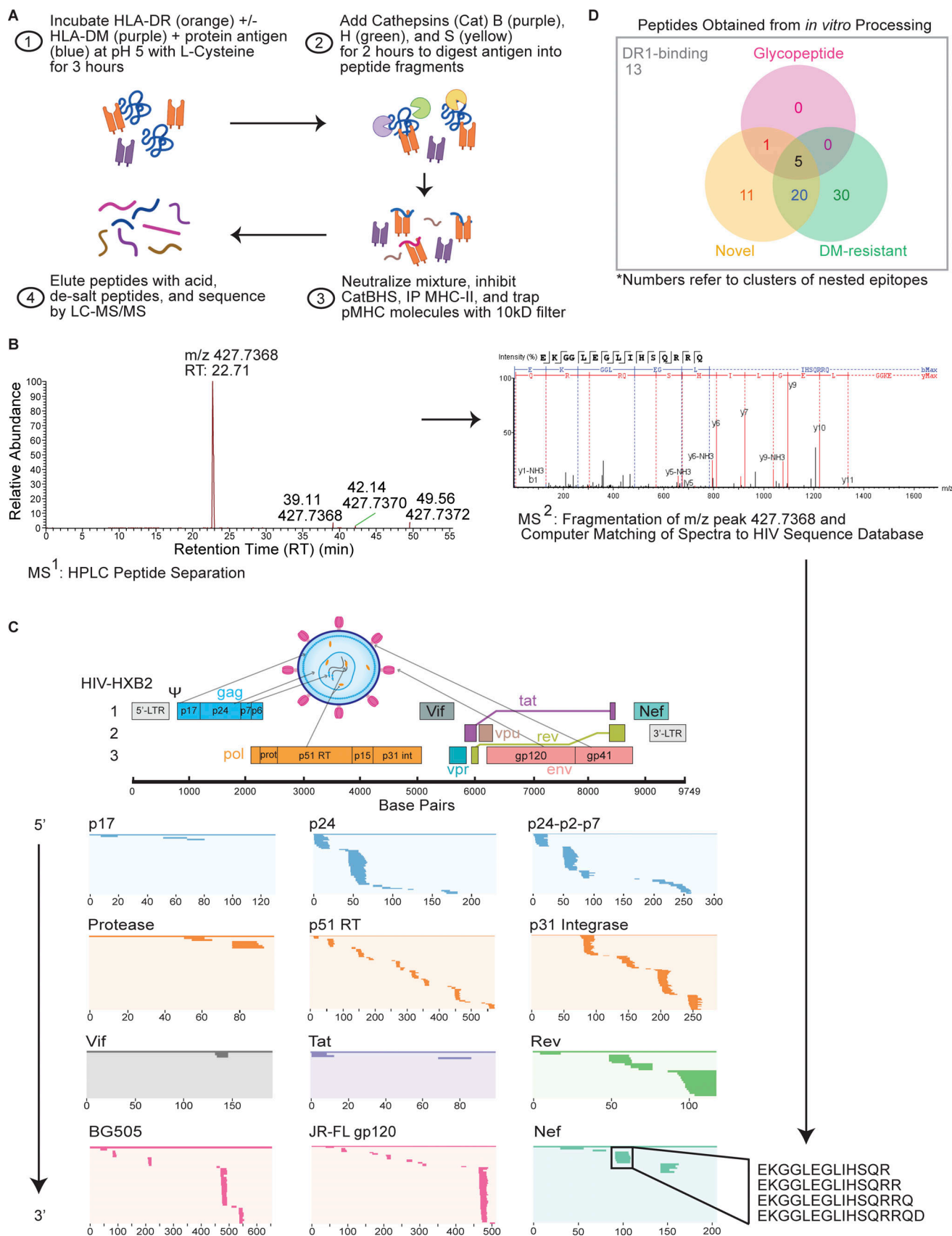


Figure 1. **Cell-free antigen processing system identifies immunodominant epitopes in HIV-1 proteome.** (A) Cell-free processing workflow. HIV-1 protein antigens were incubated with HLA-DR and +/- DM in reducing conditions and low pH for 3 h before the addition of cathepsins B, H, and S (CatBHS) for 2 h. The

solution was neutralized, cathepsins inhibited, and DR1 was immunoprecipitated. Peptides were eluted and sequences identified by LC-MS/MS. **(B)** Extracted base peak chromatograph representing a peptide from LC-MS ( $MS^1$ ) of Nef cell-free-derived epitopes. Further fragmentation of the peptide ( $MS^2$ ) resulted in individual b and y ions corresponding to amino acids for the Nef-EKG<sub>93-108</sub> epitope (Table S4). **(C)** The HIV-1 genome with its three reading frames is shown, along with the near-full proteome that was subjected to cell-free processing. Locations of epitopes from in vitro processing in all conditions (+/- DM) are shown as lines within the overall protein sequence and listed in 5' to 3' order. Shown for Nef within the black box is a cluster of epitopes corresponding to Nef-EKG<sub>93-108</sub>, a hot spot obtained from cell-free processing. **(D)** Venn diagram showing number of epitope clusters identified through cell-free processing of the HIV-1 proteome. Novel epitopes exclude those previously reported to induce memory responses in HIV+ individuals or vaccine recipients (LANL, 2018) and have <60% overlap with a literature epitope (Table S4). "13" indicates the number of epitope clusters that bind to DR1 only in the absence of DM and are not novel epitopes or glycopeptides. Data in C and D represent two independent experiments performed per antigen tested in single determinations due to antigen availability and the quantity required per assay.

(Myr-MA) yielded three DM-resistant epitopes (Fig. 2 A), while processing of p24-p2-p7 (Newman et al., 2004) and p24 yielded seven and three clusters of DM-resistant epitopes, respectively (Fig. 1 C and Fig. 2, B and C). Processing of Pol proteins also produced several epitope clusters (Fig. 2, D-F). INT processing revealed nested sets of epitopes (Fig. S1 C) that resided within four discrete locations (Fig. 1 C and Fig. 2 F). Of the accessory proteins Vif, Tat, Rev, and Nef, individual epitope hot spots were also observed (Fig. 1 C and Fig. 3, A-D). Cell-free processing of the transcriptional activator Tat (Rice, 2017) yielded epitopes near the N' and C' termini (Fig. 3 B). Rev yielded four DM-resistant epitope clusters, including a predominant epitope cluster of Rev-SPQ<sub>99-116</sub> from the C-terminus (Fig. 1 C and Fig. 3 C). Cell-free processing of Nef led to four main epitope clusters, with a single DM-resistant epitope (Fig. 1 C and Fig. 3 D). Finally, cell-free processing of gp120 and gp140 proteins from HIV-1 also yielded clear hot spots (Fig. 3, E and F).

In addition to hot spots, we observed similarities in epitopes identified in HIV-1 polyproteins and individual subunits of those polyproteins, for instance with p24-p2-p7 and p24. Peptide spectral matches from MS data provided information on relative abundance of peptides within the sample (Zybailov et al., 2005). For both p24-p2-p7 and p24, two of the most abundant epitopes from cell-free processing were the overlapping epitope cluster Gag-QNY<sub>130-151</sub>/Gag-PIV<sub>133-150</sub> (Fig. 2, B and C) as well as Gag-EGA<sub>177-199</sub> (Fig. 4, A and B). This pattern of shared epitope selection between individual subunits and polyproteins was also observed with our analysis of gp140 and gp120. For both gp140 and gp120, the three epitopes with the greatest abundance were Env-EHF<sub>91-103</sub>, Env-YCA<sub>217-227</sub>, and Env-SEL<sub>481-499</sub> (Fig. 4, C and D). In vitro processing of gp120 from a different HIV-1 isolate (LAV) yielded the same three dominant epitopes (data not shown). The finding that gp120 and trimeric gp140 yielded overlapping epitopes suggests that the denaturing environment of the MIIC sufficiently exposes most gp120/gp140 high-affinity DR1 binding sites regardless of original tertiary structure and allows the trimer form to behave similarly to destabilized monomers.

The location of epitopes in certain hot spots suggested a structural etiology. To understand whether structural features of a protein antigen led to epitope selection from in vitro processing (Landry, 1997), we analyzed the solvent-accessible surface area (ASA) and thermal stability of the HIV-1 protein antigens as inferred from crystal structures (Metz et al., 2016). Protein thermal instability is likely critical for epitope dominance: unstable regions unfold first and become available to

antigen-processing machinery, leading to increased abundance of certain epitopes. To examine protein folding stability (hereafter referred to as simply protein stability), we used a thermodynamic analysis of the ensemble of possible partially unfolded states (the COREX/BEST [Biology using Ensemble-based Structural Thermodynamics] algorithm), which has been validated by Hydrogen/Deuterium exchange (Hilser and Freire, 1996; Hilser et al., 1998, 2006; Pan et al., 2000; Whitten et al., 2005) as well as nuclear magnetic resonance (NMR)-monitored acid denaturation and cold denaturation of proteins (Babu et al., 2004; Hilser and Freire, 1996; Liu et al., 2012; Whitten et al., 2006; see Materials and methods). Using this algorithm, we could predict which regions of a protein structure are less stable and more likely to unfold.

### Structurally unstable regions of HIV-1 proteins predict epitope dominance

Structural analysis was performed for HIV-1 protein antigens with available crystal structures (Table S1). For small, monomeric proteins for which the COREX algorithm was designed, epitopes obtained by cell-free processing were located in regions of low stability (Fig. 5 and Fig. S2, A-F). For example, the abundant Gag-QNY<sub>130-151</sub>/Gag-PIV<sub>133-150</sub> epitope is contained within the  $\beta$ -hairpin loop at the N-terminus of p24 (Cortines et al., 2011). Protein termini may be more available to bind to DR1 for subsequent processing. Indeed, Gag-PIV<sub>133-150</sub> contained significantly higher accessibility and lower stability than the rest of the protein (Fig. 5, A and B). Notably, only the C-terminal portion of the Gag-PIV<sub>133-150</sub> epitope was resolved from the 1E6J structure, which may reflect this region's more dynamic nature (Fig. 5 C). A different p24 structure containing the full epitope sequence provided the same conclusions (Fig. 5 B, right). Thus, structural features of Gag-PIV<sub>133-150</sub> may promote its presentation.

Structural analysis also provided insights on epitope locations within HIV-1 accessory proteins obtained from cell-free processing. Processing of Vif yielded the DR1-binding epitope Vif-CEY<sub>133-146</sub> (Fig. 3 A), which had high accessibility and low stability (Fig. 5 D). Processing of the transcriptional activator Tat (Rice, 2017) also revealed a novel N-terminal epitope Tat-MEP<sub>1-12</sub>, which had high accessibility and low stability (Fig. 5 D). Cell-free processing of Nef in the presence of DM led to convergence on a single epitope (Nef-EKG<sub>93-108</sub>; Fig. 3 D), also from a region with high accessibility and low stability (Fig. 5 E). The most abundant epitope from cell-free processing of Rev, C-terminal Rev-SPQ<sub>99-116</sub> (Fig. 3 C), was in



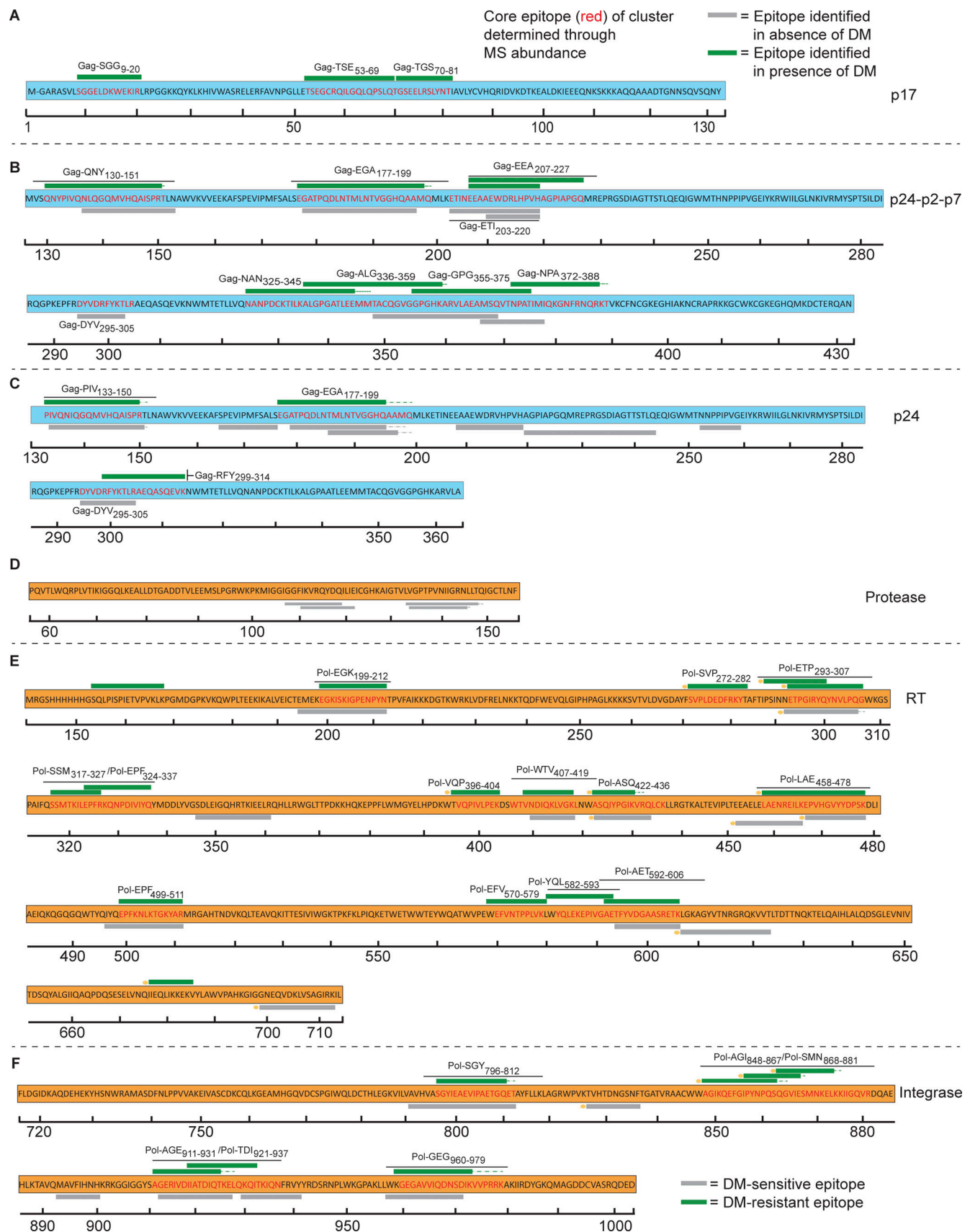


Figure 2. **Cell-free processing of Gag and Pol proteins reveal epitope hot spots.** Epitopes from the HIV-1 Gag and Pol proteins identified by LC-MS/MS from cell-free processing are shown in the form of epitope maps, with epitopes highlighted across the Gag and Pol proteins. **(A–C)** Maps for Gag proteins

include (A) p17 (where M- indicates myristoylation of the first Gly residue), (B) p24-p2-p7, and (C) p24. **(D-F)** Epitope maps for Pol proteins include (D) protease, (E) RT, and (F) INT. Green bars indicate epitopes obtained both in the presence and absence of DM (DM-resistant); gray bars indicate epitopes obtained only in the absence of DM (DM-sensitive). Hatched lines indicate additional residues (i.e., “ragged edges”) at the ends of epitopes that were observed (see below, Fig. 4, A and B). For each epitope cluster, the core epitope was defined using the peptide with the greatest number of PSMs (see below, Fig. 4, A and B). Novel epitopes (<60% overlap with existing 2018 LANL Database epitopes, see Table S4) are indicated with gold circles. Epitope maps in A–F represent two independent experiments performed per antigen tested in single determinations.

a region absent from the Protein Data Bank (PDB) structure, potentially reflecting a more disordered conformation. Altogether, in vitro processing of accessory proteins generally led to selection of epitopes with high accessibility and low stability.

COREX was less predictive of stability patterns for larger proteins, such as RT, a heterodimer comprised of a 66 kD subunit (p66) and 51 kD subunit (p51), and gp140 (Fig. S2, B–E). 60% of epitopes from RT were solvent accessible and 56.7% had low stability constants (Fig. S2, B and C). However, when all epitopes from cell-free processing were analyzed in aggregate, two-thirds of the epitopes obtained were located in regions of lower stability or higher accessibility (Fig. 5 F), and there was a statistically significant inverse correlation between accessibility and stability ( $r = -0.57$ ,  $P < 0.0001$ ). Thus, dominant epitopes identified by cell-free processing were associated with regions of lower stability and higher accessibility, with almost half of the epitopes having both these characteristics.

#### DM influences epitope diversity and relative abundance

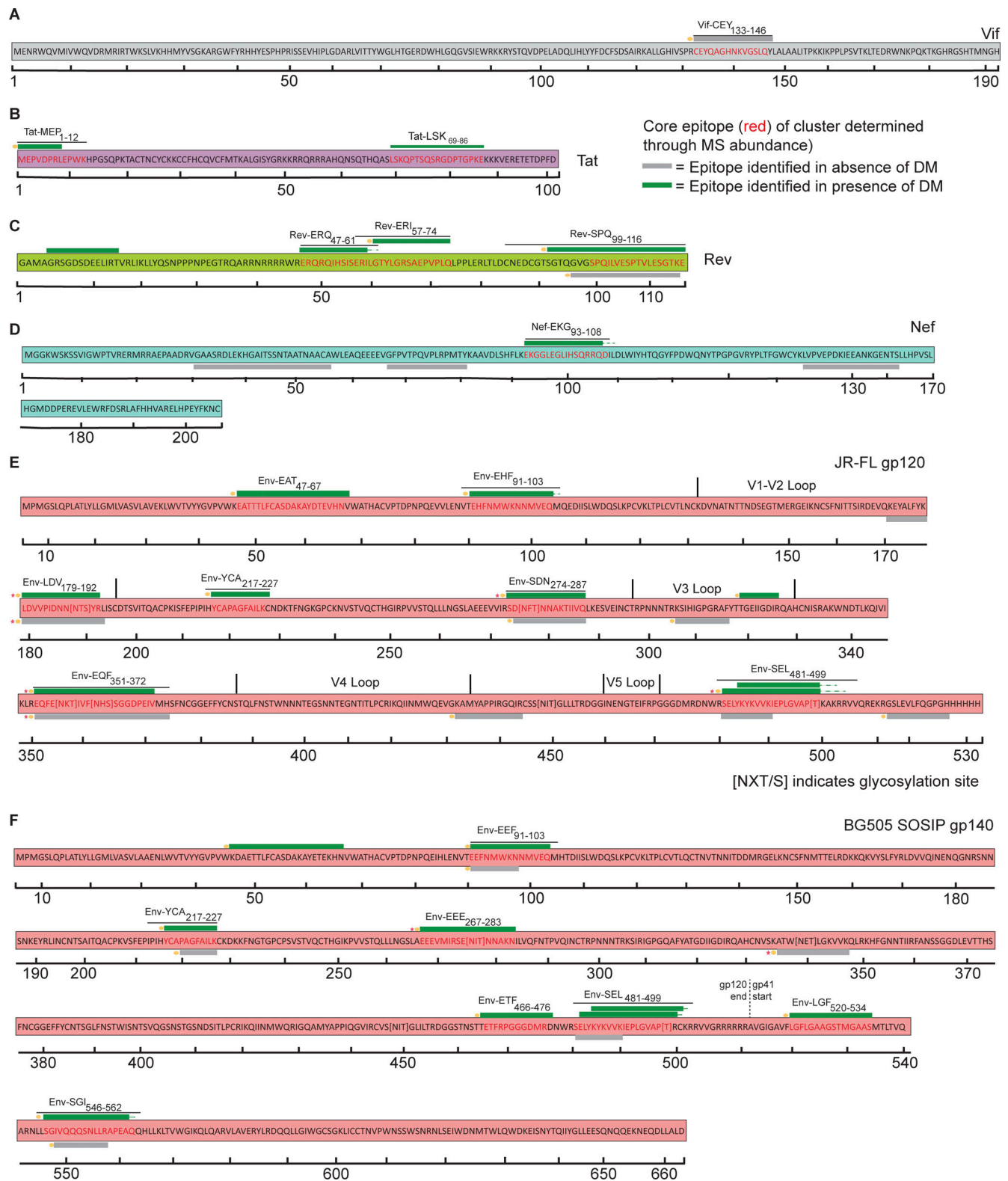
As stated above, cell-free processing experiments were performed with and without DM, as resistance to DM-mediated dissociation predicts epitope dominance (Hartman et al., 2010; Yin et al., 2012; Fig. 1, A and B). Indeed, we found that the inclusion of DM narrowed the diversity of peptides identified from cell-free processing, in some cases to one epitope. In vitro processing of p24-p2-p7 in the presence of DM reduced the number of peptide spectra from 532 to 267 sequences and revealed Gag-QNY<sub>130-151</sub>/Gag-PIV<sub>133-150</sub> as the most abundant epitope cluster (Fig. 6 A). For in vitro processing of p24 alone (without competition from p2 and p7), Gag-PIV<sub>133-150</sub> and Gag-EGA<sub>177-199</sub> again dominated peptide spectra (Fig. 6 B). In this case, there was an even more dramatic shift in abundance from the Gag-EGA to the Gag-QNY/PIV epitope cluster in the presence of DM. The ability of the Gag-QNY/PIV epitope cluster to withstand DM-mediated dissociation reflects the sequence of QNYPIVQNLQGQMVQAISPRT and the biochemical nature of DM resistance, which requires an epitope to contain a large hydrophobic residue to fill the P1 pocket of DR1 (Chou and Sadegh-Nasseri, 2000; Stern et al., 1994). The Ile, Val, and Leu contained within Gag-QNY/PIV are candidate P1 pocket residues. Importantly, the core epitope required for binding to DR1 (QNYPIVQNLQGQMVQAISPRT; the underlined portion of the epitope corresponds to the minimal core epitope required to bind to the peptide-binding groove of DR1, as determined by Harcourt et al., 1998) was contained within nearly all DM-resistant peptides from the Gag-QNY<sub>130-151</sub>/Gag-PIV<sub>133-150</sub> cluster (Fig. 4, A and B), and PIV-QNLGQMVHQAISPRL and QGQMVQAISPRTL bind with high affinity to DR1 (Harcourt et al., 1998; Wilson et al., 2001). Thus,

the peptide sequence, high affinity for DR1, DM resistance of Gag-QNY<sub>130-151</sub>/Gag-PIV<sub>133-150</sub>, and the location of the epitope in a lower stability region as discussed above, may promote its presentation.

Remarkably, the addition of DM to in vitro processing of Nef also narrowed the peptide repertoire, in this case to a single epitope (Nef-EKG<sub>93-108</sub>; Fig. 6 C). This epitope has been shown to induce CD4<sup>+</sup> T cell responses in HIV<sup>+</sup> individuals (Table S2). In another example, DM narrowed the diversity of peptides obtained from cell-free processing of gp140 and gp120: the epitope Env-SEL<sub>481-499</sub> represented 50 and 75% of DM-resistant peptides derived from cell-free processing of gp140 (Fig. 6 D) and gp120 (Fig. 6 E), respectively. This peptide is also a known CD4<sup>+</sup> T cell epitope (Table S2) and binds to DR1 with an IC<sub>50</sub> of 4 nM (Fonseca et al., 2006; Fig. 4 D and Fig. 6 D, top). Its MS abundance may reflect an optimal DR1 core sequence (YKVVKIEPL) that would favor capture of the antigen by DR1 and relatively higher accessibility (Fig. S2 F), but not necessarily lower stability. This could be a case where optimal sequences override less-than-optimal structures. In another example highlighting the role of DM in influencing epitope selection, three DM-resistant epitopes from integrase were most abundant by MS (Pol-SSM<sub>317-327</sub>/Pol-EPF<sub>324-337</sub>, Pol-AET<sub>592-606</sub>, and Pol-WTV<sub>407-419</sub>; Fig. S1 C); these epitopes have previously been shown to elicit CD4<sup>+</sup> T cell responses from people living with HIV (PLWH; Table S2).

Unexpectedly, in some instances, in vitro processing produced only DM-sensitive epitopes. This was the case for PR and Vif (Fig. 2 D and Fig. 3 A). This finding may reflect fewer optimal DR1-binding registers within these relatively smaller protein sequences. In other cases, DM did not appear to narrow peptide species diversity. Cell-free processing of RT produced 19 epitope clusters (Fig. 6 F). Most clusters contained at least one DM-resistant peptide, and several DM-resistant epitopes that are more abundant by MS (Pol-SSM<sub>317-327</sub>/Pol-EPF<sub>324-337</sub>, Pol-AET<sub>592-606</sub>, and Pol-WTV<sub>407-419</sub>) have previously been shown to elicit CD4<sup>+</sup> T cell responses from PLWH (Table S2). Notably, two DM-resistant epitopes identified in our assay (Pol-ETP<sub>293-307</sub> and Pol-EEA<sub>452-466</sub>/Pol-LAE<sub>458-478</sub>) are novel (Table S2). Given the high numbers of DM-resistant DR1-restricted RT epitopes observed, it is surprising that more RT-specific CD4<sup>+</sup> T cell responses have not been documented. This may reflect low levels of RT expression or an antigen-independent factor (such as a lower frequency of naive T cells recognizing Pol proteins; Campion et al., 2014).

Overall, these results show that while the extent of DM-mediated dissociation can differ for different protein antigens, the presence of DM in the antigen-processing compartment clearly affects epitope abundance, and that epitope hierarchy is



**Figure 3. Cell-free processing of HIV-1 accessory proteins and Env yields epitope hot spots and reveals overlapping epitopes. (A–D)** Epitope maps of (A) Vif, (B) Tat, (C) Rev, and (D) Nef peptides obtained from in vitro cell-free processing. **(E and F)** Epitope maps of peptides identified from cell-free processing of (E) gp120 (JR-FL strain) and (F) gp140 (BG505 SOSIP). As above, green bars indicate epitopes obtained both in the presence and absence of DM (DM-resistant); gray bars indicate epitopes obtained only in the absence of DM (DM-sensitive). Novel epitopes are indicated by gold circles. Pink stars indicate epitopes containing an N- or O-linked glycosyl moiety. Epitope maps in A–F represent two independent experiments performed per antigen tested in single determinations.



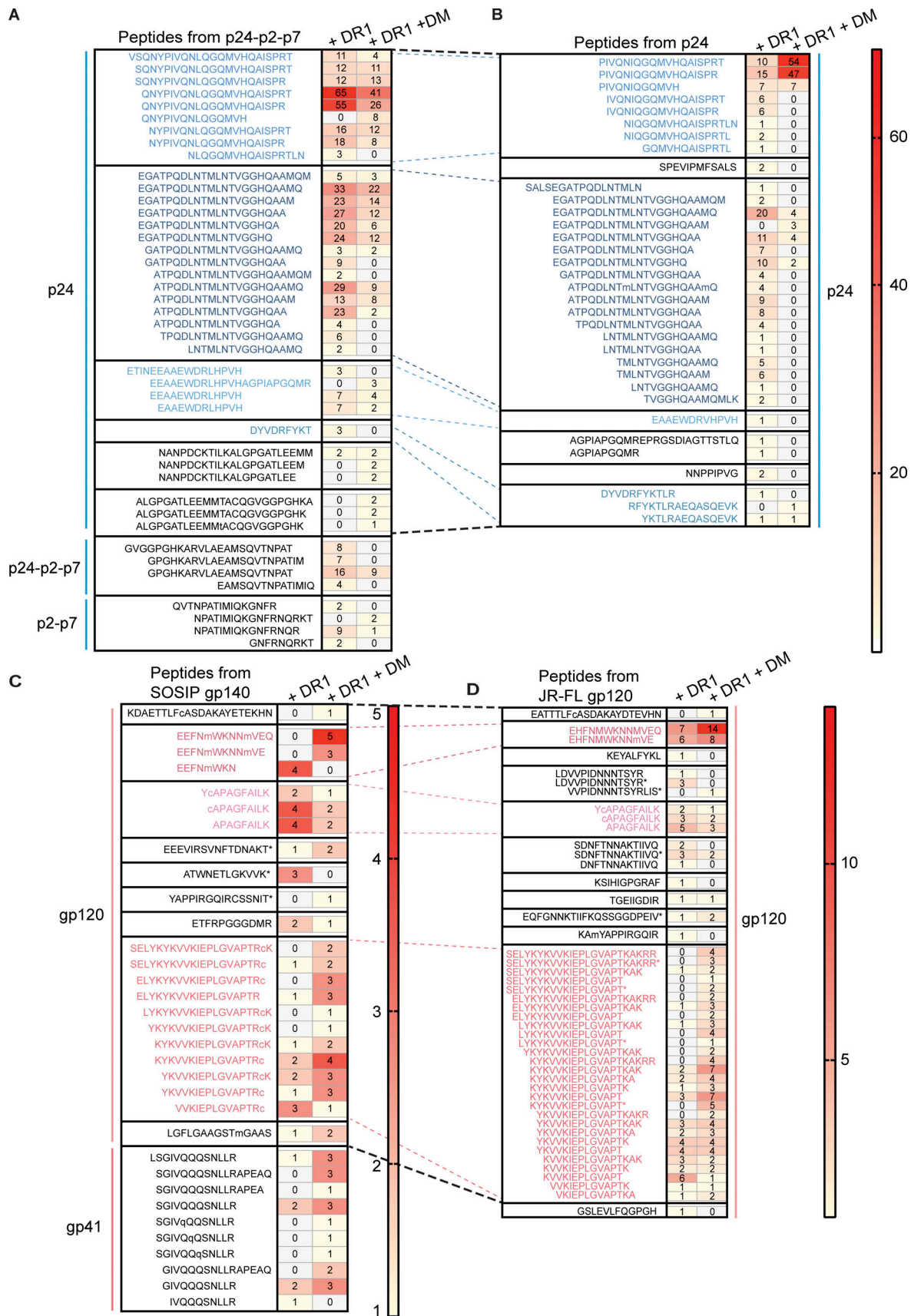


Figure 4. **Cell-free processing reveals similarities in epitopes from HIV-1 polyproteins versus individual subunits.** (A and B) Heat maps displaying PSMs identified at >95% probability from in vitro processing of (A) p24-p2-p7 (eight clusters) or (B) p24 (seven clusters) in the presence of DR1 +/- DM. Peptides



within solid lines indicate a cluster or nested set of epitopes. Peptide clusters were defined based on shared start and end residues, as well as the extent of overlap between the P1 and P9 anchor residues for DR1. p24-p2-p7 contains a VSQNY extension from p17 at the 5' end. Slight sequence differences from the p24 portion of p24-p2-p7 (utilizing NL4.3 lab strain, which contains the NY5 sequence for Gag) compared to p24 alone (HXB2, an infectious molecular clone of LAV) can be observed in certain epitopes, such as the Leu to Ile mutation in Gag-PIV: PIVQNIQGQMVHQAI SPR. **(C and D)** Heat maps showing PSM differences from cell-free processing of (C) gp140 (BG505 SOSIP) and (D) gp120 (JR-FL) in the presence of DR1 +/- DM. Black hashed lines in A–D indicate where the individual protein (i.e., p24) derives from the polyprotein (i.e., p24-p2-p7). Light blue (Gag) or pink (Env) hashed lines indicate common epitopes shared between the individual protein (B, D) and the polyprotein (A, C). Asterisks in C and D indicate glycopeptides.

first governed at the level of antigen processing (Sadegh-Nasseri and Kim, 2019). More abundantly presented epitopes may induce stronger T cell responses, leading to immune escape. Indeed, the abundant Gag-PIV<sub>133–150</sub> epitope discussed above was among the least conserved of p24 epitopes (Fig. S3 A), with observed mutations affecting the Ile, Ala, and Ile residues (Fig. S3 B) that may affect TCR recognition rather than binding to DR1 (Harcourt et al., 1998). Thus, the resistance of an epitope to DM-mediated dissociation and the expected increase in epitope density on the cell surface may influence the number of T cells responding to the presented pMHC-II.

#### Cell-free system exposes inefficiency in processing of protective epitopes

Integrating protein structure and DM resistance provided insights into a surprising observation: the low abundance of the C-terminal Gag-RFY<sub>299–314</sub> (RFYKTLRAEQASQEV) epitope from cell-free processing (Fig. 6, A and B). This epitope largely overlaps a previously described, highly conserved immunodominant epitope Gag293 (FRDYVDRFYKTLRAEQASQE), which elicits responses in >50% of PLWH (Kaufmann et al., 2004) and is associated with control of viral replication (Ranasinghe et al., 2012; Vingert et al., 2010). Unexpectedly, we observed few peptide spectral matches (PSMs) in this region of Gag: one PSM for Gag-DYV<sub>295–305</sub> (DYDRFYKT) from p24-p2-p7 that was DM-sensitive (Fig. 4 A) and two PSMs from p24 alone (RFYKTLRAEQASQEV and YKTLRAEQASQEV) that were DM-resistant (Fig. 4 B). We did not observe these latter two peptides from p24-p2-p7 (Fig. 4 A), possibly reflecting competition from p2/p7 epitopes and/or the different conformation of the polyprotein. The low PSM number is unlikely to be due to our detection limit as we detected femtomole quantities of these peptides in spiked-in experiments (Fig. S3, C and D). As YVDRFYKTLRAEQASQEV had an experimental IC<sub>50</sub> value of 5 nM for binding to DR1 (Ranasinghe et al., 2013), poor peptide binding to DR1 is unlikely to be responsible. We conclude that although Gag-RFY<sub>299–314</sub> is immunodominant from the perspective of the T cell response, it is processed inefficiently from p24. Gag-RFY<sub>299–314</sub> is highly accessible (Fig. 5 A) but located in a highly stable protein region, in contrast to most p24 epitopes that are located in regions of low to average stability (Fig. 5 B). Notably, another low-abundance epitope associated with spontaneous control (AFSPEVIPMFALSSEGA; the underlined part of this previously published epitope refers to the specific sequence that was observed by MS from our cell-free analysis; Fig. 4 B; Ranasinghe et al., 2012) was also located in a high-stability region (Fig. 5 B). Overall, p24 appears to have highly conserved epitopes associated with immune control of viral replication that may not

be well presented, and improved immunogen design may allow for enhanced protective T cell responses.

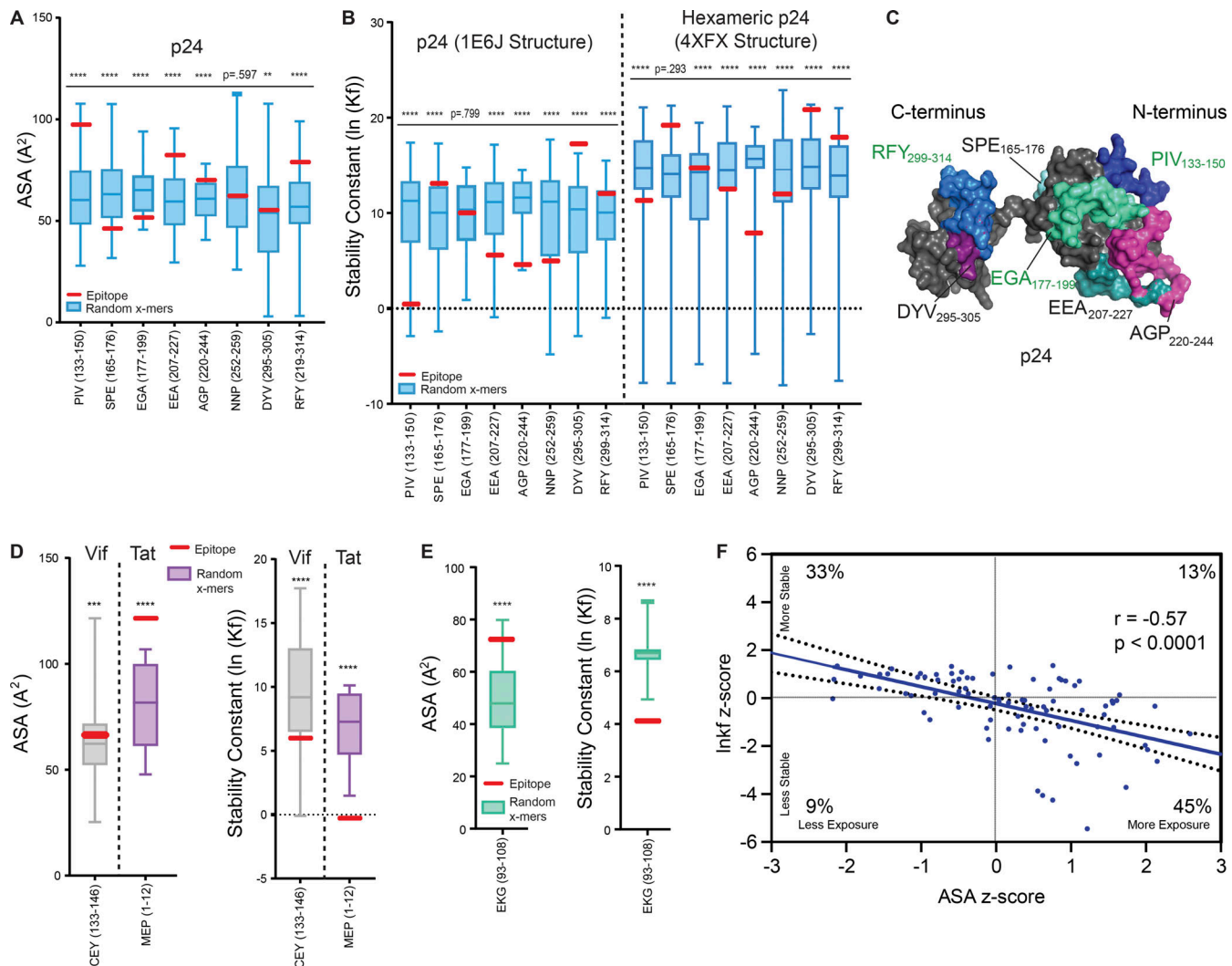
#### In vitro processing yields epitopes not previously described from T cell response data

Encouragingly, all HIV-1 proteins yielded epitopes via cell-free processing that corresponded to epitopes previously reported to induce memory CD4<sup>+</sup> T cell responses (Table S2). As hypothesized, however, we found that our cell-free processing system revealed novel epitopes as well, most apparent from Pol, Env, and accessory proteins. We identified two novel DM-resistant epitopes from RT (Pol-ETP<sub>293–307</sub> and Pol-EEA<sub>452–466</sub>/Pol-LAE<sub>458–478</sub>; Table S2). From INT, the cell-free processing system also yielded two novel epitopes—Pol-AGI<sub>848–867</sub> and Pol-SMN<sub>868–881</sub> (Table S2)—which map to regions of high accessibility, low stability, and low mutation frequency, suggesting potential utility in HIV-1 vaccine design.

Analyzing the epitopes identified from Env cell-free processing also revealed novel epitopes. Of the three most abundant DM-resistant epitopes (Env-EHF<sub>91–103</sub>, Env-YCA<sub>217–227</sub>, and Env-SEL<sub>481–499</sub>), only Env-SEL<sub>481–499</sub> had been previously described (Table S2). Processing of the BG505 SOSIP trimer also yielded a novel gp120 epitope, Env-ETF<sub>466–476</sub>, and two novel DM-resistant gp41 epitopes: Env-SGI<sub>546–562</sub> and Env-LGF<sub>520–534</sub> (Fig. 4 C), the latter epitope overlapping with the gp41 fusion peptide (Fig. 3 F). In addition to these peptides, we wondered whether we could be missing potential glycosylated epitopes by our analysis. Glycans comprise ~50% of HIV-1 gp120 by weight (Shen et al., 2014). The gp120/gp140 epitopes revealed by in vitro processing were primarily located in the vicinity of the gp120/gp41 interface or the CD4 binding site (Fig. 7, A and B). Epitopes were not found in variable loops, potentially reflecting the enrichment in N-linked glycosylation sites in these loops (Fig. 3, E and F).

To uncover potential glycosylated epitopes that may be overlooked by traditional LC-MS/MS, we analyzed our MS data on DR1-bound peptides from in vitro processing using GPQuest, an algorithm that evaluates glycopeptides (Sun et al., 2016; Toghi Eshghi et al., 2015). We identified four glycopeptide families from gp120 and two from gp140/BG505 (Table S3), which to our knowledge represent the first HIV-1 glycopeptides that can bind to a human DR molecule. As the gp120 and trimeric gp140 analyzed here were produced in 293T cells, their O- or N-linked glycosylation moieties should reflect mammalian glycosylation patterns. These results are intriguing in light of recent findings that a gp120 glycopeptide epitope bound to murine I-A and I-E elicited CD4<sup>+</sup> T cell responses in immunized mice in a glycan-dependent manner (Sun et al., 2020).

A glycan moiety located within an epitope's core binding register may affect T cell recognition. In that sense, the



**Figure 5. Exposed or structurally unstable regions of HIV-1 proteins predict epitope dominance.** (A) Epitopes from p24 (Fig. 4 B) were analyzed for solvent accessibility (ASA) by PDB PISA algorithms from the PDB structure 1E6J. Higher values of ASA indicate higher epitope exposure. Data are displayed as box-and-whisker plots showing the epitope of interest (EOI) in red compared to a distribution of average ASA values of random x-mers spanning the entirety of the protein (derived via sliding scale analysis, see Materials and methods). These distributions exclude the EOI in red so that its stability relationship with the rest of the protein can be visualized. (B) Epitopes from p24 were analyzed for stability, expressed as average stability constants (lnKf) by the COREX/BEST algorithm, from PDB structures 1E6J and 4XFX. Data are displayed in box-and-whiskers plots as in A. Lower values of lnKf indicate lower epitope stability. (C) p24 epitopes obtained from cell-free processing in the absence (black text labels) or presence (green text labels) of DM are highlighted (PDB: 1E6J). Epitope NNP<sub>252-259</sub> located on the posterior surface of 1E6J is not shown. (D and E) Comparison of accessibility (PDB PISA, left) and stability (COREX, right) of dominant epitopes obtained from cell-free processing of Vif (D, gray), Tat (D, purple), and Nef (E) relative to a distribution of randomly generated x-mer epitopes spanning these proteins (see Methods), excluding the EOI. EOI is shown in red. Structures used for analysis in D and E are shown in Table S1. (F) Two-tailed Pearson correlation was used to analyze the relationship between z-scores of ASA (PDB PISA) versus lnKf (stability, COREX) for all peptides obtained from cell-free processing. Best fit line is shown encased in a 95% confidence interval. Quadrants dividing the data into more/less stable and more/less accessible regions of the graph were used to obtain frequencies of epitopes within each group. Data in A, B, D, and E that were normally distributed were subject to a one-sample, two-tailed t test, and non-normally distributed data were subject to a two-tailed Wilcoxon Signed Rank Test, comparing the mean (t test) or median (Wilcoxon Signed Rank) of the distribution to the mean stability of the epitope. \*\*P < 0.01; \*\*\*P < 0.001; \*\*\*\*P < 0.0001.

DM-resistant glycoepitope EEE<sub>267-283</sub> in BG505 is particularly interesting (Table S3). Glycopeptide EEE<sub>267-283</sub> is modified with a single N-acetylglucosamine (Table S3). The modified Asn does not affect peptide/DR1 binding, as it was eluted from immunoprecipitated DR1 following cell-free processing. However, the glycan is close to or within the binding register for DR1, which is likely VMIRSENIT. This epitope is located on an external unstable loop making this more accessible to the MHC-II groove (Fig. S2 F). It also overlaps the binding footprints of >15

broadly neutralizing antibodies (bNAbs) that target the CD4 binding site (Fig. 7 B). Notably, one of 10 DR1+ PLWH tested (Donor 3641) showed a memory T cell cytokine response to the glycosylated but not unglycosylated form of EEE<sub>267-283</sub> (Fig. S3 E and Fig. S4 A). Further studies will need to verify whether glycan-dependent T cell responses are observed in larger cohorts and if such responses impact humoral immunity. Aside from EEE<sub>267-283</sub>, we observed several examples of T/B cell epitope overlap at the CD4 binding site or gp120/gp41 interface (Fig. 7 B).

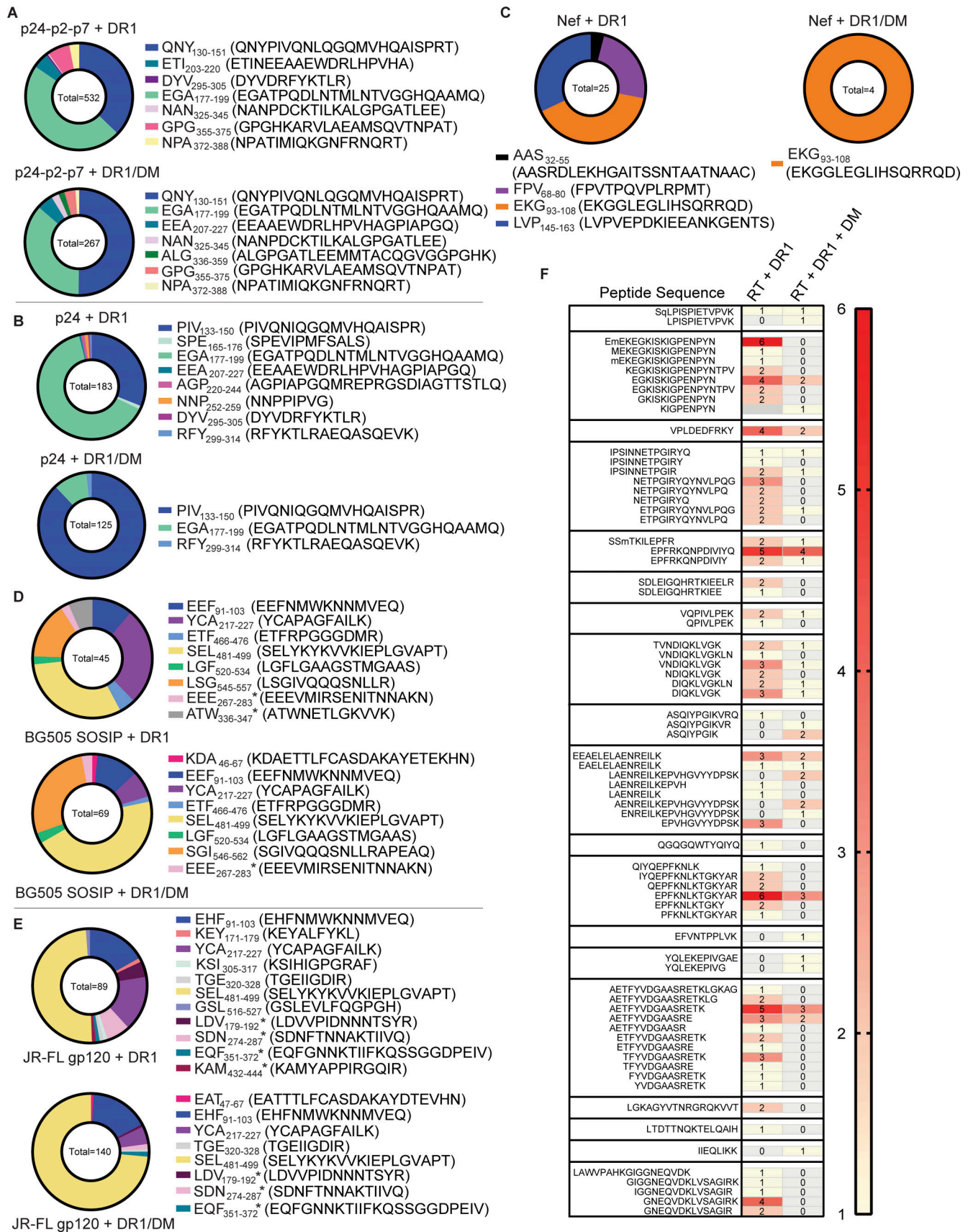


Figure 6. Cell-free processing displays influence of DM influences on epitope dominance and relative abundance. (A and B) Relative differences in abundance of various (A) p24-p2-p7 or (B) p24 epitopes are shown as differences in PSMs in the absence (above) or presence (bottom) of DM. The total number







a single MHC-II allele. We evaluated whether these peptides (Table S4) could be presented *in vivo* by analyzing CD4<sup>+</sup> T cell responses from DRI<sup>+</sup> PLWH on suppressive combination anti-retroviral therapy (cART; Table S5). Early administration of cART preserves T cell immunity (Altfeld et al., 2001; Le et al., 2013; Ndhlovu et al., 2019). Of 10 donors studied, five were treated during acute infection (<6 mo after infection; Table S5). CD4<sup>+</sup> T cell responses were measured by intracellular cytokine staining for IFN $\gamma$ , IL-2, and TNF $\alpha$  (Fig. S3 E and Fig. 8 A). As internal controls, we included a representative sample of previously published epitopes (Fig. 8 B, denoted in asterisks), as well as the irrelevant self-peptide CLIP<sub>89-105</sub> (KMRMATPLLM-QALPM) that binds to nascent DRI (Fig. 8 A). Responses were confirmed with multiple cytokines and were not observed in cells from DRI<sup>+</sup> HIV<sup>-</sup> healthy donors ( $n = 6$ ; Fig. S4 B). Responses were considered positive if they were polyfunctional (at least two or more cytokine responses; Fig. S4, C–E, and Fig. S5, A–E) and represented a greater than 2.95-fold increase in magnitude relative to the response to CLIP.

Of the donors tested, Donor 3037 displayed the greatest breadth of responses (Fig. 8, A and B). Donor 3037, who first tested positive for HIV-1 in 1979 (from a retrospective sampling of banked specimens), started on nucleoside reverse transcriptase inhibitor monotherapy in the late 1980s and on cART in 1998 (Table S5). Following development of drug-resistance mutations, a new and fully suppressive regimen was started in 2018. Despite the time interval between infection and suppressive therapy, Donor 3037 had polyfunctional responses to two of the three dominant DM-resistant epitopes identified in cell-free processing of p24: Gag-PIV<sub>133-150</sub> and Gag-RFY<sub>299-314</sub> (Fig. 8 B and Fig. S5 E), in addition to the parent epitope FRD<sub>293-312</sub> described in the literature. FRD<sub>293-312</sub>, with the core epitope of Gag-RFY<sub>299-314</sub>, has been associated with viral control (Benati et al., 2016; Ranasinghe et al., 2012; Vingert et al., 2010). This donor also had a response to the overlapping epitopes Gag-ETI<sub>203-220</sub>/Gag-EEA<sub>207-227</sub> (Fig. 2 B), which are accessible and unstable epitopes that represent the third most abundant p24 epitope from p24p2p7 processing (Fig. 4 A). Furthermore, donor 3037 showed a polyfunctional response to the novel RT epitope Pol-LAE<sub>458-478</sub> (Fig. 2 E), which was the most abundant DM-resistant RT epitope from cell-free processing (Fig. 6 F). Responses to Gag-RFY<sub>299-314</sub>, Gag-FRD<sub>293-312</sub>, Gag-ETI<sub>203-220</sub>, Gag-EEA<sub>207-227</sub>, but not Gag-PIV<sub>133-150</sub>, were confirmed independently using an IFN $\gamma$  ELISPOT (Fig. S5 F). Robust CD4<sup>+</sup> T cell responses to these epitopes may have conferred some protective immunity and prevented progression to AIDS for 20 yr (Table S2).

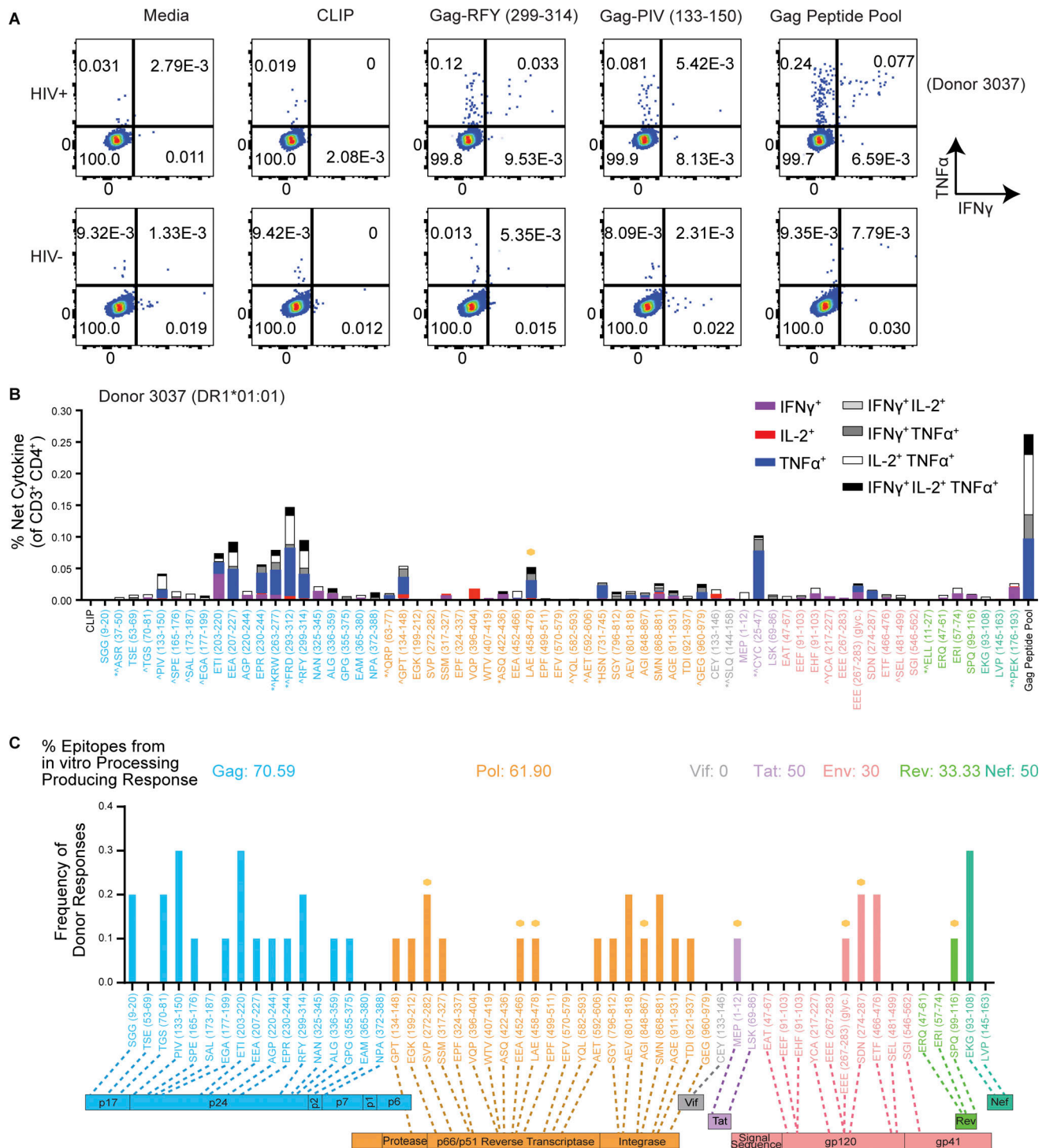
Altogether, from 10 PLWH, we observed cytokine responses to a substantial fraction of epitopes identified by *in vitro* processing: Gag, 70.6%; Pol, 61.9%; Env, 30%; Vif, 0%; Tat, 50%; Rev, 33%; and Nef, 50% (Fig. 8 C). Responses were observed to 55.3% of all DM-resistant epitopes, including novel epitopes. Most responses were specific for Gag or Pol epitopes (Fig. 9 A). Of the 56 epitopes tested, 31 produced responses in cells from PLWH, including eight of the 19 novel epitopes tested here (Fig. 9 B). Some epitopes elicited very strong responses but in only a subset of donors. Thus, a minimalist cell-free system can uncover novel epitopes that are processed and presented *in vivo*.

If efficiency of epitope processing and presentation is a determinant of T cell responses, epitope abundance from cell-free processing may be reflected in the frequency of DRI-restricted responses to that epitope. The Gag-PIV<sub>133-150</sub> epitope, which was one of the most frequently targeted by PLWH in our cohort (30%, Fig. 8 C), displayed the greatest abundance by peptide spectra (50.8 and 88% of DM-resistant epitopes from p24p2p7 and p24, respectively; Fig. 6, A and B), consistent with its structural characteristics (Fig. 5, A and B) and mutational profile (Fig. S3, A and B). Nef-EKG<sub>93-108</sub>, the most abundant DM-resistant Nef epitope by MS (Fig. 6 C), also induced CD4<sup>+</sup> T cell responses in 30% of HIV<sup>+</sup> individuals (Fig. 8 C). Overall, we observed a modest but significant correlation between peptide frequency from *in vitro* processing and CD4<sup>+</sup> T cell response frequency ( $r = 0.34$ ,  $P = 0.04$ ; Fig. 9 C).

## Discussion

Our cell-free antigen processing system interrogates how structural features of an antigen and components of the MHC-II antigen processing pathway allow regions of an antigen to be efficiently processed and emerge as dominant. This contrasts with most epitope prediction algorithms that rely on peptide affinity for an MHC allele or T cell responses to overlapping peptides. Prediction algorithms based on peptide/MHC affinity or peptide-elution data (Abelin et al., 2017; Chen et al., 2017; Jurtz et al., 2017) fail to mimic environmental conditions that create epitope hierarchies, leading to variable success in predicting MHC-II epitopes. T cell responses to overlapping peptides used at non-physiologic micromolar concentrations may omit specific registers that are optimal for MHC-II binding (Godkin et al., 2001; Lovitch et al., 2006) and cannot account for posttranslationally modified epitopes. Additionally, studies relying on IFN $\gamma$  ELISPOT assays overlook responses from individuals who have a limited CD4<sup>+</sup> IFN $\gamma$  response. Most importantly, as highlighted by Reinherz and colleagues, this “reverse immunology” approach only identifies previously recognized epitopes and may miss epitopes that could provide optimal control of infection (Keskin et al., 2015). This is especially relevant for HIV-1, where protective CD4<sup>+</sup> T cell epitopes have been documented, but most individuals progress to chronic infection.

To overcome these limitations, we subjected nearly the entire HIV-1 proteome to our reductionist cell-free antigen processing system. We identified novel epitopes that could be targeted in future vaccine studies. Examples include several conserved DM-resistant Pol epitopes. CD4<sup>+</sup> T cells specific for internal viral proteins can provide help to B cells in making antibodies to Env, as long as both are present in the viral particle internalized by the B cell (Milich et al., 1987; Russell and Liew, 1979). Targeting highly conserved intracellular epitopes from Gag and Pol proteins, rather than the more mutable Env, may prove beneficial for vaccine strategies seeking to induce robust neutralizing antibodies. One could envision a vaccine utilizing a viral vector containing Gag, followed by a heterotypic vector containing Pol, that engages T cell responses to both proteins and provides more breadth than Gag alone (Arunachalam et al., 2020; Liu et al., 2009).



**Figure 8. Epitopes from cell-free processing induce cytokine responses in DR1\*01:01 HIV<sup>+</sup> individuals.** (A) Representative flow cytometry plots of IFN $\gamma$ <sup>+</sup>TNF $\alpha$ <sup>+</sup> release from CD3<sup>+</sup>CD4<sup>+</sup> T cells after 16 h of PBMC incubation with Gag-RFY<sub>299-314</sub> and Gag-PIV<sub>133-150</sub> in a DR1\*01:01 HIV<sup>+</sup> individual (Donor 3037) compared to a DR1\*01:01 HIV<sup>-</sup> donor. Media, CLIP (irrelevant peptide), and Gag peptide pool controls are shown for comparison. (B) Representative single-, double-, and triple-positive (see legend) responses from the CD4<sup>+</sup> T cells of Donor 3037 from ex vivo PBMC stimulation with 65 peptides across the HIV-1 proteome. Single-positive responses indicate cells that produced either IL-2, IFN $\gamma$ , or TNF $\alpha$ , double-positive responses indicate cells that produced two of the three cytokines assessed, and triple-positive responses indicate cells producing all three cytokines. Asterisks denote peptides from the literature, while the remaining 56 peptides were identified from in vitro processing and selected for testing in this screen (Table S4). ^ indicates epitopes from the literature known to be restricted by HLA DR1\*01:01. Gold circles indicate novel epitopes identified from cell-free processing that induced a response. Activation after stimulation was measured in single determinations. (C) Frequency of HIV<sup>+</sup> donor responses to 56 of the epitopes obtained from cell-free processing as measured by IL-2, IFN $\gamma$ , or TNF $\alpha$  in positivity compared with CLIP within each donor. Responses were considered positive if they were polyfunctional and at least two of the cytokines measured represented a >2.95-fold increase in the magnitude of response relative to CLIP. The percent of epitopes from in vitro processing that produced a response is listed in the corresponding color. Responses to novel epitopes are indicated with a gold circle. Data in C were obtained from  $n = 10$  HIV<sup>+</sup>

donors. Activation after stimulation was measured in single determinations due to the large number of cell-free-derived epitopes tested and the cell input required for testing polyfunctional cytokine responses by flow cytometry.

Indeed, recent preventative vaccine studies in the simian immunodeficiency virus model suggest a role for help from CD4<sup>+</sup> T cells specific for intracellular viral proteins (Casimiro et al., 2005; Liu et al., 2009). Additionally, HIV<sup>+</sup> individuals who developed neutralizing antibodies and spontaneously controlled HIV-1 viral loads to <2,000 copies/ml for >1 yr in the absence of cART had stronger CD4<sup>+</sup> T cell responses to Gag than to gp120 (Ranasinghe et al., 2015).

In addition to novel Pol epitopes, we identified glycopeptide epitopes, highlighting the utility of unbiased, antigen-processing-based epitope discovery. While Env sequence variation certainly plays a role, the failure of overlapping peptide methods to detect glycopeptide responses may partly explain why fewer responses to Env have been documented compared to Gag or Nef (Kaufmann et al., 2004; Ranasinghe et al., 2012), despite the fact that Env can readily enter the exogenous and endogenous pathways of MHC-II processing following internalization (Byland et al., 2007; Callahan et al., 1993; LaBranche et al., 1995). We show here that HIV-1 Env epitopes that are glycosylated can bind to DR1, withstand DM-mediated displacement, and be presented to CD4<sup>+</sup> T cells. Thus, some Env-derived epitopes may be glycopeptides not assayed by standard methods. Measuring responses to these epitopes may allow better correlations between antibody titers and Env-specific CD4<sup>+</sup> T cell responses in vaccine studies (Pauthner et al., 2019; Sanders et al., 2015).

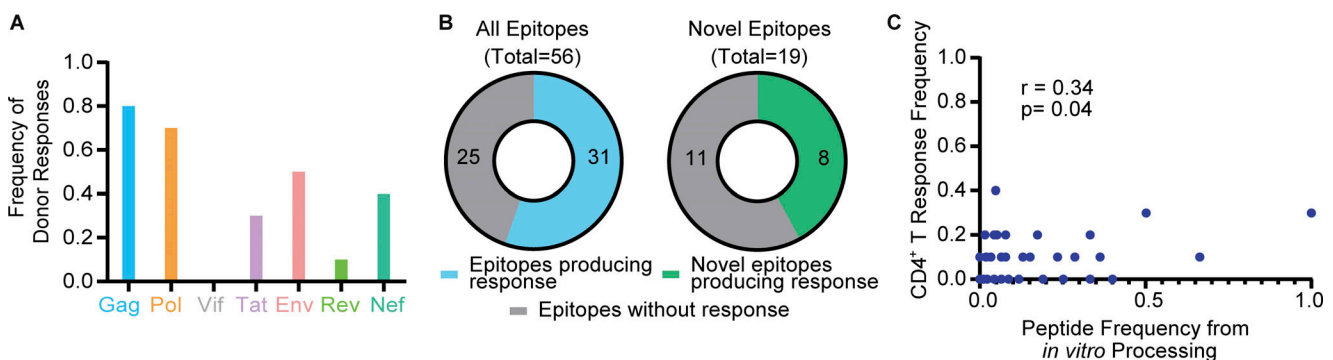
We made several observations regarding factors influencing immunodominance. We observed (a) epitope hot spots and similarities in epitopes from HIV polypeptides versus individual subunits; (b) the location of epitopes from HIV-1 proteins of known structure (p24 and accessory proteins) map to regions of low protein stability and high solvent accessibility; (c) DM influences epitope abundance; and (d) epitope selection patterns of specific HIV-1 proteins suggest that differences in processing/presentation efficiency influence T cell responses. Thus, structural features and binding to DR1 in the presence of DM allow

prediction of epitope selection and opportunities for enhanced HIV-1 vaccine strategies.

### Cell-free processing highlights protein stability and epitope location in epitope dominance

Our cell-free processing system provides insights into antigen structure that predispose regions for capture by DR1 and subsequent processing and presentation. Most epitopes identified by cell-free processing—in particular for smaller proteins such as p24, Tat, Nef, and Vif—corresponded to regions of higher accessibility and lower protein stability. Such regions have an increased chance of unfolding, permitting capture by DR1. This relationship was less pronounced with larger, multi-domain proteins such as RT and Env, perhaps reflecting the increased inaccessibility of epitopes from these proteins in solution as well as computational limitations. Additionally, while certain regions may initially be less accessible and have higher stability, binding to DR1 in the MIIC compartment may facilitate protein unfolding, allowing previously “stable” regions to become unstable and available for processing (Sadeh-Nasseri and Kim, 2019). Likewise, in low pH and denaturing conditions, localized changes in protein folding and allosteric interactions during protein oligomerization may facilitate the exposure of previously stable regions, allowing for their capture and presentation by DR1. In the low pH and denaturing conditions utilized in our cell-free system, DR1 binding sites could be exposed regardless of original tertiary structure for larger proteins. Overall, our results show that more accessible and less regions of HIV-1 protein antigens tend to be more efficiently processed.

Several identified epitopes overlapped with B cell epitopes. T/B cell epitope overlap has been shown to either boost or suppress T cell responses to epitopes contained within an antibody footprint, as antibody binding can protect the epitope from degradation upon virion endocytosis or sterically block the processing of the epitope (Jaume et al., 2002; Simitsek et al.,



**Figure 9. Characteristics of CD4<sup>+</sup> T cell responses from DR1\*01:01<sup>+</sup> HIV<sup>+</sup> individuals to epitopes identified from cell-free processing.** (A) Frequency of HIV<sup>+</sup> donor responses to individual HIV-1 antigens are shown. (B) Number of epitopes producing a CD4<sup>+</sup> T cell response of the 56 total and 19 novel epitopes tested in ex vivo stimulation experiments are shown. (C) Pearson correlation between peptide frequency from in vitro processing and donor CD4<sup>+</sup> T cell response frequency across PLWH is shown. Peptide frequency was assessed by dividing PSMs for an epitope by the total MS spectra for a particular protein. *r* and *P* values are shown. Data in A–C were obtained from *n* = 10 HIV<sup>+</sup> donors. Activation after stimulation was measured in single determinations.



1995; Watts and Lanzavecchia, 1993). These findings may provide insights into T cell immunodominance for current bNAb trials seeking to enhance virus-specific T cell responses (Niessl et al., 2020).

### Epitope abundance in the presence of DM provides clues for improved vaccine design

The presence of DM affected the relative abundance of certain epitopes, and we observed a significant correlation between relative peptide abundance by LC-MS/MS and frequency of CD4<sup>+</sup> T cell responses in HIV<sup>+</sup> individuals. Prior studies have shown that greater surface density of agonist pMHC-II both enhances T cell activation (Korb et al., 1999; Mirshahidi et al., 2004; Mirshahidi et al., 2001) and reduces the duration of the T cell priming phase (Henrickson et al., 2008; Mempel et al., 2004). DM-resistant epitopes will be presented at higher cell-surface density and induce stronger T cell activation, irrespective of naive TCR repertoire and TCR affinities, simply due to their enhanced antigen processing efficiency. It is interesting to also consider antigen density in the context of T follicular helper cell (Tfh) interactions with B cells. Tfh regulates the number of cell cycles that B cells undergo in the dark zone of the germinal center reaction, in a manner proportional to the number of cognate pMHC-II presented by those B cells (Gitlin et al., 2014). Thus, abundantly presented HIV-1 epitopes may drive Tfh responses that promote enhanced somatic hypermutation and the induction of neutralizing antibodies or even bNAbs in HIV<sup>+</sup> individuals. Examining CD4<sup>+</sup> T cell responses to abundant versus less abundantly presented epitopes may provide insights into optimal sequential vaccination regimens that induce bNAbs.

Some candidate HIV-1 vaccines use viral vectors to deliver multiple HIV-1 antigens in addition to Env (Liu et al., 2009). However, our results show protein stability and epitope competition can influence epitope selection. For example, the highly conserved Gag-RFY<sub>299-314</sub> epitope, associated with viral control (Ranasinghe et al., 2012), was not observed from the Gag polyprotein processing but was observed at low abundance from processing of the p24 subunit alone. Low PSMs of Gag-RFY<sub>299-314</sub> may reflect its location within a stable region of p24, which may undergo a delay in unfolding and be kinetically outcompeted by epitopes from less stable regions. Elite controllers may have developed compensatory mechanisms to sense lower levels of protective epitopes, including enrichment for public TCRs that have a high affinity for the extended Gag-293 epitope (Benati et al., 2016; Vingert et al., 2010). However, inducing responses to protective epitopes in non-controller individuals may be affected by epitope processing efficiency. If Gag-RFY<sub>299-314</sub> epitope is inefficiently processed, as our results suggest, then including the whole Gag polyprotein may actually reduce the presentation of this conserved epitope. In vaccine strategies geared toward inducing CD4<sup>+</sup> T cell responses to highly conserved epitopes, competition from other epitopes within the same protein and inefficiencies in antigen processing should be considered.

While most identified epitopes induced responses by CD4<sup>+</sup> T cells from PLWH, some did not. This may indicate the following: (1) the epitope was not presented in vivo, potentially due

to competition from other HIV-1 proteins; (2) the epitope was presented and primed CD4<sup>+</sup> T cells that became exhausted or anergic; (3) the individual's naive precursor and/or memory CD4<sup>+</sup> T cell frequency for the pMHC-II was limited (Campion et al., 2014; Moon et al., 2007); (4) the protein from which the epitope is derived is expressed at a relatively lower level, leading to less peptide presentation; (5) the viral variant that primed these CD4<sup>+</sup> T cells has sequence differences in the epitope; or (6) our cohort size may not have captured the full spectrum of responses to all the HIV-1 epitopes tested here. Future studies with larger cohort sizes may improve the chance of detecting positive responses to such epitopes. Larger cohorts may also provide insights into whether antigen-dependent factors predict epitopes that may be dually recognized by CD4<sup>+</sup> T cells and very rare MHC-II restricted CD8<sup>+</sup> T cells that have been observed in humans (Ranasinghe et al., 2016). Finally, studies with larger cohorts of PLWH would help establish whether the correlation we observed between epitope abundance and CD4<sup>+</sup> T cell response frequencies is enhanced, which would have significant implications for understanding mechanisms of immunodominance and improved vaccine development.

While this study provides several insights into factors influencing epitope selection, there are a few limitations. A potential limitation of this approach is the preferential activation and HIV infection of HIV-specific CD4<sup>+</sup> T cells (Douek et al., 2002). Recent studies have shown that some clones of HIV-infected, HIV-specific CD4<sup>+</sup> T cells persist and proliferate in PLWH on cART (Collora et al., 2022; Simonetti et al., 2021). Nevertheless, it is important to note that in the setting of preventative vaccines, the use of novel epitopes identified with this system may provide better T cell help for a neutralizing antibody response that would prevent any infection of CD4<sup>+</sup> T cells. In the context of a therapeutic vaccine, the use of epitopes identified with this system may allow the generation of a broader CD4<sup>+</sup> T cell response including epitopes that were not targeted by the initial response to infection in PLWH. If the immunization is done in the setting of cART, no new cells will be infected.

Our use of ex vivo peptide stimulation experiments to validate the epitopes obtained from cell-free processing, while commonly used to assess for CD4<sup>+</sup> T cell memory responses to pathogen-derived peptides, does not incontrovertibly demonstrate in vivo processing and presentation of these antigens. Tracking of tetramer-positive CD4<sup>+</sup> T cells after the onset of acute infection and functionally characterizing isolated tetramer-positive cells for proliferation and cytokine release in response to peptide or whole protein antigens would provide additional evidence to support in vivo presentation during natural infection. Additionally, since the cell-free system obviates the need for identifying the nature of the APC, it is unclear whether infected CD4<sup>+</sup> T cells present these peptides directly or whether DCs indirectly present these epitopes during the course of natural infection (Addison et al., 2022). Finally, the cell-free processing system in this study generates an HIV-1 peptidome for a single MHC-II allele, DR1\*01:01, and as such is not completely generalizable. However, results from these analyses can provide insights into overall epitope selection trends across other HLA alleles.



## Conclusions and future studies

Altogether, these results show that epitope selection based on resistance to DM-mediated dissociation and the structure of the antigen influence epitope selection leading to memory CD4<sup>+</sup> T cell responses in PLWH. The relative abundance of epitopes presented by DR1 in the presence of DM was reflected by the frequencies of observed CD4<sup>+</sup> T cell responses. Future studies investigating the relative contribution of each APC to epitope-specific CD4<sup>+</sup> T cell responses may be achievable in humanized mouse studies. Additionally, future studies that incorporate multiple HIV-1 proteins in the cell-free assay could reveal which epitopes dominate when the input antigen is a whole virion (Kim et al., 2017). In vitro processing of proteins produced from mosaic vaccines or viral-vectored vaccines could also provide information on the presentation of specific HIV-1 epitopes associated with protection. Given the central role of CD4<sup>+</sup> T cells in coordinating adaptive immunity, understanding epitope selection of vaccine antigens is invaluable not only for HIV-1 but also for other challenging pathogens such as malaria and *Mycobacterium tuberculosis*, and for understanding differential vaccine responsiveness to SARS-CoV-2.

## Materials and methods

### Experimental model and subject details

#### Human subjects

Functional assays were performed using peripheral blood mononuclear cells (PBMCs) from 10 DR1<sup>+</sup> (HLA-DR1\*01:01) HIV<sup>+</sup> individuals from the University of California, San Francisco (UCSF) Study on the Consequences of the Protease-Inhibitor Era (SCOPE) cohort (see Table S5 for additional details on study participants). All study subjects provided written informed consent before participation in the study, and the study was approved by the UCSF Institutional Review Board. Deidentified PBMCs from six DR1<sup>+</sup> (HLA-DR1\*01:01) HIV<sup>+</sup> individuals were obtained via leukapheresis samples (STEMCELL) as controls.

#### Cell lines

Sf9 and Hi5 cells were used to produce recombinant DR1 and DM. Cells were grown as previously described (Kim et al., 2014; Narayan et al., 2007) in ISFM media (Cleveland et al., 2014; Inlow et al., 1989) at a density of 0.3–0.5 million cells/ml at 27°C, with regular passaging every 3–4 d.

### Method details

#### Production of recombinant proteins

Soluble HLA-DR1\*01:01 and DM were produced as described (Kim et al., 2014; Narayan et al., 2007). Baculovirus DNA (BaculoGold; PharMingen) and transfer vectors carrying DR  $\alpha$ - and  $\beta$ -chains were transfected together into Sf9 insect cells to produce recombinant viruses. Recombinant viruses were passaged three times before being used to infect High Five cells in ISFM media (Cleveland et al., 2014; Inlow et al., 1989). DR1 proteins were purified from culture supernatants using immunoaffinity chromatography with a monoclonal antibody L243 to DR1 (purified from HB-55 hybridoma; American Type Culture Collection). Soluble DM was also expressed by High Five cells

transduced with baculovirus containing the extracellular domains of genes encoding the  $\alpha$ - and  $\beta$ -chains of human DM. The truncated DM  $\alpha$ - and  $\beta$ -chains were modified to contain the FLAG epitope (DYKDDDDK) and c-Myc epitope (EQKLISEEDL), respectively, at their C-termini. DM protein was purified from culture supernatants with a monoclonal antibody to M2 (anti-FLAG) agarose resin (Sigma-Aldrich) and eluted with 5 mg/ml FLAG peptide (Sigma-Aldrich) in tris-buffered saline. DM was further concentrated and buffer exchanged into citric phosphate buffer, pH 6, with 0.5% wt/vol sodium azide and stored in aliquots at  $-80^{\circ}\text{C}$ .

#### Cell-free processing assay

Cell-free processing experiments were conducted as described (Hartman et al., 2010; Kim et al., 2014). Specifically, on day 0, 20  $\mu\text{M}$  DR1 was preincubated with 400  $\mu\text{M}$  HA(Y308A), which forms short-lived complexes with DR1 and generates a peptide-receptive DR1 conformation (Kim et al., 2014). After incubation overnight at 37°C, 325 pmol of peptide-receptive DR1:Y308A was added to 750 pmol protein antigen in the presence or absence of 162.5 pmol of DM, together in citrate-phosphate buffer (pH 5–5.2) with 6 mM L-Cysteine. This mixture was incubated at 37°C for 3 h. After this time, a cathepsin digestion mixture consisting of 90 pmols Cathepsin B (bovine spleen, Sigma-Aldrich), 90 pmols cathepsin H (human liver, Calbiochem), and 38 pmols cathepsin S (human recombinant protein produced in *Escherichia coli*, Calbiochem), as well as 4 mM EDTA were added to the reaction for an additional 2 h. At the end of cathepsin digestion, the pH of the assay was adjusted to 7.4 with equal volumes of 2 M sodium dibasic buffer and 1X PBS, and 10  $\mu\text{M}$  iodoacetamide was added to inhibit cathepsin activity. Peptide-bound DR1 molecules were immunoprecipitated with HLA-DR-specific mAb (L243)-conjugated Sepharose beads for 1 h at 4°C. DR1-bound beads were washed with PBS and water, and pMHC-II molecules were eluted from the antibody-conjugated beads with mild acid elution (0.1% TFA). Peptides were subsequently eluted from DR1 with 1% TFA and 40% MeOH/1% TFA and physically separated from MHC molecules using a 10 kD MWCO filter (Millipore) before being lyophilized dry.

#### Hydrophilic interaction LC (HILIC) cleanup

Peptides were cleaned for MS by HILIC using columns (HILIC TopTips) containing poly (2-sulfoethylaspartamide)-silica membranes (Alpert and Andrews, 1988). Lyophilized peptides were resuspended in 85% acetonitrile (ACN) in 15 mM ammonium formate ( $\text{NH}_4\text{HCO}_2$ ) at room temperature. HILIC TopTips were conditioned by washing with 0, 2, 10, and 85% ACN/ $\text{NH}_4\text{HCO}_2$ . Sample was added to the column in 10  $\mu\text{l}$  increments. Columns were washed with 85% ACN/ $\text{NH}_4\text{HCO}_2$ , and bound peptides were eluted first in 10%, then 2%, and finally 0% ACN/ $\text{NH}_4\text{HCO}_2$ . Peptides were lyophilized and identified by LC-MS/MS.

#### MS

Peptides were separated via reverse-phase chromatography with an Easy nLC 1000 (Thermo Fisher Scientific), using a gradient of 2–90% ACN/formic acid over 60 min at 300 nl per

min on a C18 column packed with MAGIC AQ C18 at 3  $\mu$ m, 100 Å (MICHROM Bioresources, Inc.). Eluting peptides were sprayed onto the nano-LC-Q-Exactive Plus Orbitrap (Thermo Fisher Scientific) through a 10- $\mu$ m integrated emitter tip at 2.2 kV. Survey scans (full mass spectra) were acquired on the Orbitrap within 350–1,800 D m/z using the data-dependent Top 10 method with dynamic exclusion of 10 s. Precursor ions were individually isolated with 1.6 D and fragmented (MS/MS) using high energy collisional dissociation activation collision energy 28. Precursor and fragment ions were analyzed at a resolution of 140,000/35,000 at 200 D. Automatic gain control target 3 $\times$ e6 max IT 60 ms and automatic gain control target 1 $\times$ e5, mx IT250 ms for parent and fragment ions, respectively.

### MS data analysis

Tandem MS2 mass spectra were analyzed by Proteome Discoverer (v1.4 Thermo Fisher Scientific) in three ways, using 3Nodes: common, Xtract (spectra are extracted, charge state deconvoluted, and deisotoped using Xtract option, at resolution 105 K at 200 D) and MS2 Processor. MS/MS spectra from 3Nodes were analyzed with Mascot v2.5.1 Matrix Science ([www.matrixscience.com](http://www.matrixscience.com)) using a Custom Database (2015RefSeq\_72r\_human with added client proteins database) consisting of sequences from all HIV-1 proteins tested (Table S6) as well as molecules of the in vitro processing system (MHC-II, DM, and cathepsins), and a concatenated decoy database, specifying the following search parameters: “no enzyme,” precursor mass tolerance of 8 ppm, fragment mass tolerance of 0.02 D, and variable modifications (cysteine carbamidomethylation, methionine oxidation, and other custom modifications). “No enzyme” was used to detect peptides generated by cleavage after any residue. Mascot “.dat” files were compiled in Scaffold. Scaffold Viewer Software (version Scaffold\_4.8.9, Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Scaffold uses the Protein Prophet algorithm to assign probabilities for protein identification (Nesvizhskii et al., 2003) and the PeptideProphet or LDFR algorithm to assign probabilities for peptide identification. Peptide identifications used for analysis were those that could be established at 95% probability to achieve a false discovery rate (FDR) <1% by the Peptide Prophet algorithm (Keller et al., 2002) with Scaffold delta-mass correction. Proteins used for analysis were those that could be identified at 99% probability to achieve an FDR of <1% and contained more than one identified peptide (Welsh et al., 2020). Spectral counting via Scaffold was done using the Total Spectra method, which uses the sum of all spectra associated with a specific protein within a sample, and also includes spectra shared with other proteins. Peptides identified from PSMs after in vitro processing were grouped into clusters based on shared start and end residues and their extent of overlap. DR1 contains a nine-residue binding groove, between the P1 and P9 anchor positions, so we considered the overlap of residues within this region. If two peptides had different anchor residues (i.e., Tyr, Phe, Pro, Ile, Leu) or lacked a nine-residue overlap within the DR1 binding groove, they were considered distinct epitopes. For each epitope cluster, the edges of each candidate epitope were defined using the peptide with the greatest number of PSMs,

which is referred to as the core epitope in the relevant figures. Epitopes that were considered DM sensitive were those which were detected in the +DR1 only condition by the above criteria but could not be detected at an appropriate significance in the +DR1/+DM condition (see Fig. S1 B for an example).

### Glycopeptide identification

Glycopeptides were identified using the GPQuest software (Toghi Eshghi et al., 2015). Briefly, LC-MS/MS raw files were converted to mzML files using the msconvert tool in ProteoWizard with peak picking function. In GPQuest search, mass tolerance of MS1 and MS2 levels were 10 and 20 ppm, respectively. In-house N-linked glycan database with 277 compositions and O-linked glycan database with 83 compositions from Functional Glycomics Gateway (CFG, <http://www.functionalglycomics.org/fg/>) were used. The peptide database included 2,042 human N-linked glycopeptides and tryptic peptides of gp120 with two missed cleavages. MS/MS spectra were filtered to have at least three oxonium ions, and an oxonium ion at 204 m/z was mandatory. Information on oxonium ions was used to predict the type of glycosylation that facilitated data interpretation. A modified Morpheus score was calculated using singly charged –b and –y peptide fragment ions and peptide + glycan fragment ions, and epitopes with a score >6.16 was used to identify hits. FDR was calculated as described previously. The identified glycopeptides were manually checked to ensure the quality of MS/MS spectra for identification.

### Accessibility and structural stability analysis

Solvent ASA were obtained by inputting PDB structures in Table S1 into the PDB PISA tool. The ASA of each residue for the relevant protein chain in an interface could be extracted. The stability constants for each residue were also obtained from the respective PDB structures for each protein using the COREX/BEST web program (Hilser and Freire, 1996; Vertrees et al., 2005). Specifically, the algorithm models equilibrium conformational fluctuations of a protein to generate an ensemble of microstates, capturing the partial unfolding observed from hydrogen exchange and NMR relaxation experiments (Bai et al., 1995; Hilser and Freire, 1997). The COREX algorithm apportions the protein into “folding windows” that are overlaid onto the high-resolution crystal structure to generate a collection of differently folded states of the protein. The thermodynamic contribution of each Boltzmann-weighted state in this ensemble to the overall thermodynamic properties of the protein can be used to obtain a probability of local unfolding for different regions of the protein structure.

To calculate the free energy of each microstate, the calorimetrically parameterized enthalpy and heat capacity of the state are obtained from the change in solvent-accessible area from protein unfolding from the crystal structure (Murphy and Freire, 1992). The total entropy of the state is determined by the sum of the calorimetrically parameterized solvation entropy (due to the change in solvent accessibility) and the weighted conformational entropy (estimated from molecular dynamics simulation; D’Aquino et al., 1996; Lee et al., 1994). These thermodynamic properties can be used to determine the free energy

of the microstate and thus, the statistical weight of that state in contributing to the overall ensemble. From this, the algorithm determines stability constants, essentially an equilibrium constant for folding, which provide the summed probability of states in which a residue is in a folded conformation over the number of states in which residues are in an unfolded conformation. These stability constants are the primary data plotted in Fig. 5. Residues that are more stable will be folded in most high-probability states while residues with low stability constants will be unfolded in most high-probability states. This method of calculating protein stability based on high-resolution structural data has been validated experimentally by hydrogen/deuterium exchange (Hilser and Freire, 1996; Hilser et al., 1998, 2006; Pan et al., 2000; Whitten et al., 2005) as well as NMR-monitored acid denaturation and cold denaturation of proteins (Babu et al., 2004; Hilser and Freire, 1996; Liu et al., 2012; Whitten et al., 2006).

For our analyses, a window size of eight residues, minimum window size of four, and 10,000 microstates per partition were selected for generating the ensemble for each protein. Exceptions were for the PDB structures 4ZMJ (BG505 SOSIP) and 1HMY (unliganded p66 and p51 subunits, analyzed separately), where a window size of 10 and 9 were used, respectively, to accommodate the larger size of the proteins. These represent an increased degree of sampling as compared to the defaults for the Monte Carlo option provided by the COREX/BEST server. All other unspecified parameters were default values. Thus, each protein ensemble was composed of  $\sim 8 \times 10,000 = 80,000$  partially unfolded states.

### Sliding scale analyses

A sliding scale of random epitopes was generated across the length of the respective protein to quantify whether an epitope's solvent accessibility or epitope stability differed significantly from random epitopes throughout the protein (see Table S8 for an example). Each epitope of interest contains accessibility and stability values associated with every residue. These values can be averaged to give an accessibility or stability score for the epitope. To compare this score to randomly generated epitopes across the protein, we customized a sliding scale analysis for each epitope. If the epitope was a 15-mer, then a sliding scale of random 15-mers spanning the protein—excluding any 15-mers that touched the epitope of interest—was generated. Since each random 15-mer has accessibility and stability constants associated with each residue, it can also be assigned an accessibility and stability score by taking the average of these values.

The accessibility and stability scores for the entire set of random 15-mers can then be averaged and compared to the accessibility/stability score for the 15-mer epitope of interest. Thus, the random set was an internal control for each epitope, specific for the epitope's complete protein. If the distribution of scores from the random 15-mers was normal, then a one-sample *t* test was used to quantify statistical significance between the mean of the distribution versus the value of the specific epitope. If the distribution of average accessibilities or stabilities was non-normal, then the Wilcoxon Signed Rank tests was used to quantify statistical significance between the median of the distribution

compared to the value of the specific epitope. A *P* value of  $<0.05$  was considered significant. Overall, this strategy provided a quantification for understanding how extreme a particular accessibility or stability score of an epitope was from the sample mean or median. This same analysis was conducted for target epitopes of different lengths (17-mers, 18-mers) for each of the HIV proteins. Distributions of the random sliding scale epitopes that excluded the epitope of interest were visualized to ensure that the overall distribution was unaltered before performing statistical analyses.

### Mismatch analysis

To determine the level of conservation of residues within an epitope, the HXB2 protein sequence was compared to a collection of sequences from patients in the Los Alamos National Laboratory (LANL) HIV sequence database to assign a number of mismatches counted to each residue within a protein. The protein sequences downloaded from the database were of HIV subtype B, from plasma, and only included sequences that were 0% non-ACGT. Only one sequence per patient was included.

In addition to mismatches at the single residue level, we assessed the prevalence of full epitopes by aligning full epitope sequences to collections of patient sequences from LANL as described. Epitopes were aligned using an adapted Boyer-Moore alignment algorithm. Epitope prevalence was assessed in the context of zero permitted mismatches or one permitted mismatch between the epitope sequence and the collection of patient sequences. Data are shown as the percentage of patient sequences which contain the epitope.

### Peptide synthesis

All peptides (Elim Biopharmaceuticals and JPT) were reconstituted at 5 mM in a solution of 50% dimethylformamide (Sigma-Aldrich) and 50% diethyl pyrocarbonate-treated water.

### PBMC isolation

HLA-typed PBMCs from DR1<sup>+</sup> HIV<sup>-</sup> healthy donors (Stem Cell) and DR1<sup>+</sup> HIV<sup>+</sup> participants from the UCSF SCOPE cohort were isolated using a Ficoll gradient and cryopreserved. Cryopreserved PBMCs were thawed and rested for 2–5 h at 37°C in R10 medium (RPMI 1640 with 10% FBS and 1% penicillin/streptomycin) before use in stimulation assays.

### Ex vivo stimulations

Rested cells were washed and plated into a 96-well U-bottom plate at a range of  $5 \times 10^5$  cells per well, depending on cell recovery. Cells were stimulated with 5  $\mu$ M of the peptides listed in Table S4 for 16–18 h in the presence of 10  $\mu$ M T20 (National Institutes of Health AIDS Reagent Program) and 1  $\mu$ g/ml brexidin A (BD Biosciences) at 37°C and 5% CO<sub>2</sub>. Peptides selected for the screen were generally DM-resistant epitopes that were dominant from LC-MS/MS, as well as a single control peptide per HIV-1 protein from the literature that had been documented to induce CD4<sup>+</sup> T cell responses in the 2018 LANL Database. Cells treated with phytohemagglutinin (0.5  $\mu$ g/ml; Remel Inc.), HIV-1 Consensus B Gag peptide pool (1  $\mu$ g ml<sup>-1</sup> per peptide; JPT), or HIV-1 Consensus Subtype B Nef peptide pool (1  $\mu$ g ml<sup>-1</sup> per



peptide; JPT), served as positive controls, while cells stimulated with human short CLIP<sub>89-105</sub> peptide (Elim Biopharm) or unstimulated (media only) served as negative controls. After washing 1× with PBS, PBMCs were incubated in Fc block (BD Biosciences) for 15 min at 23°C to prevent non-specific binding. Cells were stained with a viability dye and surface markers (15 min, 4°C) followed by intracellular cytokine staining (ICS). ICS was performed using the fixation/permeabilization solution kit (BD Biosciences) according to the manufacturer's protocol, and BV Brilliant Stain Buffer (BD) was used during ICS at 10 µl/test. Samples were acquired on an LSRII flow cytometer (BD; see Fig. S3 for antibody staining panel). Negative controls were used together with fluorescence minus one controls to set gates for analysis with FlowJo software (Treestar).

### ELISPOT

IFN $\gamma$  ELISPOTs were performed using the ELISpot Pro: Human IFN- $\gamma$  kit (3420-2HST; Mabtech) according to the manufacturer's protocol. 250,000 PBMCs from SCOPE participant 3037 were plated per well in RPMI with 10% FBS and anti-retrovirals (tenofovir and emtricitabine). The PBMCs were cultured for 20 h with 5µM peptide (Elim Biopharm) or 1 µg/ml/peptide HIV-1 Consensus B Gag peptide pool (JPT Peptide Technologies, PM-HIV-CONB). The plates were read by a blinded independent investigator using the AID iSpot Spectrum Reader. Six replicate wells were run for each condition, and data are representative of two independent experiments. *pi* Significant difference relative to CLIP determined by one-way ANOVA with Dunnett's test for multiple comparisons, \**P* < 0.05, \*\*\*\**P* < 0.0001.

### Quantification and statistical analysis

For all antigens tested in the cell-free system, at least two independent experiments were performed, one with both DM and DR and one with only DR. For ex vivo stimulation experiments, PBMCs were assayed from *n* = 10 HIV<sup>+</sup> on suppressive cART and *n* = 6 HIV<sup>-</sup> DR1\*01:01-expressing donors. Because of the limited nature of samples (PBMCs from HIV<sup>+</sup> individuals), the large number (56) of cell-free derived epitopes being tested, and the cell input required for testing polyfunctional cytokine responses by flow cytometry, activation after stimulation was measured in single determinations. Simplified Presentation of Incredibly Complex Evaluation software was used to analyze flow cytometry data on T cell polyfunctionality as previously described (Roederer et al., 2011). The ELISPOT analysis represents data from two independent experiments. For box-and-whiskers plots depicting accessibility and stability metrics, normally distributed data were subject to a one-sample, two-tailed t-test, and non-normally distributed data were subject to a two-tailed Wilcoxon Signed Rank Test, comparing the mean (*t* test) or median (Wilcoxon Signed Rank) of the random epitope distribution to the mean ASA or stability of the epitope of interest. Statistical details of experiments can be found in the individual figure legends. Significance of all *P* values reported are as follows: \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < or = 0.0001. NS, not significant. Statistical analyses were performed in GraphPad Prism 8.0 or Matlab.

### Online supplemental material

Fig. S1 provides examples of the extracted base peak chromatographs from LC-MS/MS following cell-free processing and depicts how relative PSMs identified from LC-MS/MS can be displayed via a heat map. Fig. S2 provides accessibility and stability trends for HIV-1 proteins. Fig. S3 depicts the mutational characteristics and limit of detection for cell-free epitopes, as well as a representative gating strategy for assessing CD4<sup>+</sup> T cell responses from PLWH. Fig. S4 illustrates a glycopeptide-specific memory CD4<sup>+</sup> T cell response and the lack of responses observed in healthy donors, as well as polyfunctional CD4<sup>+</sup> T cell responses detected in three out of 10 HIV<sup>+</sup> donors. Fig. S5 depicts polyfunctional CD4<sup>+</sup> T cell responses in the remaining HIV<sup>+</sup> donors as well as IFN $\gamma$  ELISPOT results from Donor 3037. Table S1 describes the HIV-1 proteins subjected to cell-free processing. Table S2 lists epitopes obtained from cell-free processing that were reported in previous studies. Table S3 lists glycopeptides obtained from cell-free processing identified by GPQuest. Table S4 lists select peptides identified from cell-free processing. Table S5 describes the clinical characteristics of PLWH in the study. Table S6 lists sequences of proteins utilized in the cell-free processing system. Table S7 provides raw data obtained from cell-free processing (Excel). Table S8 provides an example of COREX analysis for Myr-MA (Excel). Table S9 provides the CD4<sup>+</sup> T cell responses from PLWH (shown as fold change over CLIP) in ex vivo stimulations (Excel).

### Data availability

All data supporting the findings of this study are available within the paper or the supplementary materials and from the lead contact upon request. All original code used for epitope analysis has been deposited at Zenodo and is publicly available at <https://doi.org/10.5281/zenodo.6360804>. Further information and requests for reagents generated or used in this study are available upon request from the lead contact, Scheherazade Sadegh-Nasseri ([ssadegh@jhmi.edu](mailto:ssadegh@jhmi.edu)).

### Acknowledgments

We thank the following laboratories for providing HIV-1 proteins used for cell-free processing: Dr. Michael Summers, University of Maryland, Baltimore County, Baltimore, MD, USA (Myr-MA and CA-SP1-NC proteins); Dr. Stuart Le Grice, National Cancer Institute, Frederick, MD, USA (RT); and Dr. John Mascola, Vaccine Research Center, Bethesda, MD, USA (JR-FL gp120 and BG505.664 SOSIP). The following HIV-1 proteins were obtained through the National Institutes of Health HIV Reagent Program: HXB2 p24 (ARP-13126); HXB2 Protease (ARP-11781); p66/p51 Reverse Transcriptase Protein with N-Terminal Histidine Tag (ARP-3555); NL4-3 Integrase (F185H/C280S) (ARP-9420); HXB2 Vif (ARP-11050); Nef (ARP-13342); HXB2 Rev (ARP-12707); and IIB Tat (ARP-2222). Fig. 1 A was made with <https://BioRender.com>. We thank Ms. Tricia Nilles, Dr. Hao Zhang, and Dr. Elizabeth Thompson for consulting on ex vivo stimulation experimental design and data analysis. We thank Dr. Sandra Gabelli and Ms. Stormy Ruiz for assistance with structural analysis. We thank Dr. Joel Blankson for assistance with ELISPOT



experimental design. We also thank members of the Siliciano and Sadeh-Nasseri labs for suggestions regarding data analysis and visualization, and Dr. Francesco Simonetti for careful review of the manuscript.

This work was supported by grants from National Institute of Allergy and Infectious Diseases (R01AI120634, to S. Sadeh-Nasseri), the Howard Hughes Medical Institute (R.F. Siliciano), and National Institute of Allergy and Infectious Diseases F30 AI136704 (S. Sengupta).

Author contributions: S. Sengupta designed and performed experiments, analyzed data, and wrote the manuscript. J. Zhang, M.C. Reed, and J. Yu assisted in data acquisition/analysis and manuscript preparation. A. Kim produced proteins for cell-free processing assay and advised on the details of the cell-free system. T. Boronina performed MS analysis. N.L. Board performed ELISPOT analysis. J.O. Wrabl assisted with COREX/BEST analysis, data visualization, and manuscript preparation. K. Shenderov assisted with ex vivo stimulation experiment analysis and manuscript preparation. R.A. Welsh provided samples for MS analysis. W. Yang and A.E. Timmons performed glycopeptide identification and mutation counting, respectively. R. Hoh and S. Deeks provided samples from HLA-typed HIV<sup>+</sup> individuals. R.N. Cole oversaw MS data acquisition and analysis. J.D. Siliciano assisted with data analysis and writing the manuscript. R.F. Siliciano and S. Sadeh-Nasseri supervised all studies, provided funding, contributed to experimental design, and assisted in writing the manuscript.

Disclosures: S.G. Deeks reported personal fees from Tendel during the conduct of the study. No other disclosures were reported.

Submitted: 26 September 2022

Revised: 1 February 2023

Accepted: 23 March 2023

## References

Abelin, J.G., D.B. Keskin, S. Sarkizova, C.R. Hartigan, W. Zhang, J. Sidney, J. Stevens, W. Lane, G.L. Zhang, T.M. Eisenhaure, et al. 2017. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity*. 46:315–326. <https://doi.org/10.1016/j.immuni.2017.02.007>

Addison, M.M., Ellis, G.I., Leslie, G.J., Zawadzky, N.B., Riley, J.L., Hoxie, J.A., and Eisenlohr, L.C. 2022. HIV-1-Infected CD4<sup>+</sup> T Cells Present MHC Class II-Restricted Epitope via Endogenous Processing. *J Immunol*. 209: 864–873. <https://doi.org/10.4049/jimmunol.2200145>

Alpert, A.J., and P.C. Andrews. 1988. Cation-exchange chromatography of peptides on poly(2-sulfoethyl aspartamide)-silica. *J. Chromatogr. A*. 443: 85–96. [https://doi.org/10.1016/S0021-9673\(00\)94785-X](https://doi.org/10.1016/S0021-9673(00)94785-X)

Altfield, M., E.S. Rosenberg, R. Shankarappa, J.S. Mukherjee, F.M. Hecht, R.L. Eldridge, M.M. Addo, S.H. Poon, M.N. Phillips, G.K. Robbins, et al. 2001. Cellular immune responses and viral diversity in individuals treated during acute and early HIV-1 infection. *J. Exp. Med.* 193:169–180. <https://doi.org/10.1084/jem.193.2.169>

Anders, A.K., M.J. Call, M.S.E.D. Schulze, K.D. Fowler, D.A. Schubert, N.P. Seth, E.J. Sundberg, and K.W. Wucherpfennig. 2011. HLA-DM captures partially empty HLA-DR molecules for catalyzed removal of peptide. *Nat. Immunol.* 12:54–61. <https://doi.org/10.1038/ni.1967>

Arunachalam, P.S., T.P. Charles, V. Joag, V.S. Bollimpelli, M.K.D. Scott, F. Wimmers, S.L. Burton, C.C. Labranche, C. Petitdemange, S. Gangadhara, et al. 2020. T cell-inducing vaccine durably prevents mucosal

SHIV infection even with lower neutralizing antibody titers. *Nat. Med.* 26:932–940. <https://doi.org/10.1038/s41591-020-0858-8>

Babu, C.R., V.J. Hilser, and A.J. Wand. 2004. Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation. *Nat. Struct. Mol. Biol.* 11:352–357. <https://doi.org/10.1038/nsmb739>

Bai, Y., T.R. Sosnick, L. Mayne, and S.W. Englander. 1995. Protein folding intermediates: Native-state hydrogen exchange. *Science*. 269:192–197. <https://doi.org/10.1126/science.7618079>

Benati, D., M. Galperin, O. Lambotte, S. Gras, A. Lim, M. Mukhopadhyay, A. Nouël, K.A. Campbell, B. Lemerrier, M. Claireaux, et al. 2016. Public T cell receptors confer high-avidity CD4 responses to HIV controllers. *J. Clin. Invest.* 126:2093–2108. <https://doi.org/10.1172/JCI83792>

Byland, R., P.J. Vance, J.A. Hoxie, and M. Marsh. 2007. A conserved dileucine motif mediates clathrin and AP-2-dependent endocytosis of the HIV-1 envelope protein. *Mol. Biol. Cell*. 18:414–425. <https://doi.org/10.1091/mbc.e06-06-0535>

Callahan, K.M., J.F. Rowell, M.J. Soloski, C.E. Machamer, and R.F. Siliciano. 1993. HIV-1 envelope protein is expressed on the surface of infected cells before its processing and presentation to class II-restricted T lymphocytes. *J. Immunol.* 151:2928–2942. <https://doi.org/10.4049/jimmunol.151.6.2928>

Campion, S.L., T.M. Brodie, W. Fischer, B.T. Korber, A. Rossetti, N. Goonetilleke, A.J. McMichael, and F. Sallusto. 2014. Proteome-wide analysis of HIV-specific naive and memory CD4(+) T cells in unexposed blood donors. *J. Exp. Med.* 211:1273–1280. <https://doi.org/10.1084/jem.20130555>

Casimiro, D.R., F. Wang, W.A. Schleif, X. Liang, Z.-Q. Zhang, T.W. Tobery, M.-E. Davies, A.B. McDermott, D.H. O'Connor, A. Fridman, et al. 2005. Attenuation of simian immunodeficiency virus SIVmac239 infection by prophylactic immunization with dna and recombinant adenoviral vaccine vectors expressing Gag. *J. Virol.* 79:15547–15555. <https://doi.org/10.1128/JVI.79.24.15547-15555.2005>

Castellino, F., F. Zappacosta, J.E. Coligan, and R.N. Germain. 1998. Large protein fragments as substrates for endocytic antigen capture by MHC class II molecules. *J. Immunol.* 161:4048–4057. <https://doi.org/10.4049/jimmunol.161.8.4048>

Chen, B., M. Khodadoust, N. Olsson, E. Fast, L. Wagar, C.L. Liu, M. Davis, R. Levy, J. Elias, R. Altman, et al. 2017. Maria: Accurate prediction of MHC-II peptide presentation with deep-learning and lymphoma patient MHC-II ligandome. *Blood*. 130:1486

Chou, C.L., and S. Sadeh-Nasseri. 2000. HLA-DM recognizes the flexible conformation of major histocompatibility complex class II. *J. Exp. Med.* 192:1697–1706. <https://doi.org/10.1084/jem.192.12.1697>

Cleveland, T.E., IV, J.M. McCabe, and D.J. Leahy. 2014. Detergent-solubilized Patched purified from Sf9 cells fails to interact strongly with cognate Hedgehog or Ihog homologs. *Protein Expr. Purif.* 104:92–102. <https://doi.org/10.1016/j.pep.2014.09.012>

Collora, J.A., R. Liu, D. Pinto-Santini, N. Ravindra, C. Ganoza, J.R. Lama, R. Alfaro, J. Chiarella, S. Spudich, K. Mounzer, et al. 2022. Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity*. 55:1013–1031.e7. <https://doi.org/10.1016/j.immuni.2022.03.004>

Cortines, J.R., E.B. Monroe, S. Kang, and P.E. Prevelige Jr. 2011. A retroviral chimeric capsid protein reveals the role of the N-terminal  $\beta$ -hairpin in mature core assembly. *J. Mol. Biol.* 410:641–652. <https://doi.org/10.1016/j.jmb.2011.03.052>

Douek, D.C., J.M. Brenchley, M.R. Betts, D.R. Ambrozak, B.J. Hill, Y. Okamoto, J.P. Casazza, J. Kuruppu, K. Kunstman, S. Wolinsky, et al. 2002. HIV preferentially infects HIV-specific CD4<sup>+</sup> T cells. *Nature*. 417:95–98. <https://doi.org/10.1038/417095a>

D'Aquino, J.A., J. Gómez, V.J. Hilser, K.H. Lee, L.M. Amzel, and E. Freire. 1996. The magnitude of the backbone conformational entropy change in protein folding. *Proteins*. 25:143–156. <https://doi.org/10.1002/prot.1>

Fonseca, S.G., A. Coutinho-Silva, L.A.M. Fonseca, A.C. Segurado, S.L. Moraes, H. Rodrigues, J. Hammer, E.G. Kallás, J. Sidney, A. Sette, et al. 2006. Identification of novel consensus CD4 T-cell epitopes from clade B HIV-1 whole genome that are frequently recognized by HIV-1 infected patients. *AIDS*. 20:2263–2273. <https://doi.org/10.1097/01.aids.0000253353.48331.5f>

Freed, E.O. 2015. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* 13:484–496. <https://doi.org/10.1038/nrmicro3490>

Gitlin, A.D., Z. Shulman, and M.C. Nussenzweig. 2014. Clonal selection in the germinal centre by regulated proliferation and hypermutation. *Nature*. 509:637–640. <https://doi.org/10.1038/nature13300>

- Godkin, A.J., K.J. Smith, A. Willis, M.V. Tejada-Simon, J. Zhang, T. Elliott, and A.V.S. Hill. 2001. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. *J. Immunol.* 166:6720–6727. <https://doi.org/10.4049/jimmunol.166.11.6720>
- Harcourt, G.C., S. Garrard, M.P. Davenport, A. Edwards, and R.E. Phillips. 1998. HIV-1 variation diminishes CD4 T lymphocyte recognition. *J. Exp. Med.* 188:1785–1793. <https://doi.org/10.1084/jem.188.10.1785>
- Hartman, I.Z., A. Kim, R.J. Cotter, K. Walter, S.K. Dalai, T. Boronina, W. Griffith, D.E. Lanar, R. Schwenk, U. Krzych, et al. 2010. A reductionist cell-free major histocompatibility complex class II antigen processing system identifies immunodominant epitopes. *Nat. Med.* 16:1333–1340. <https://doi.org/10.1038/nm.2248>
- Henrickson, S.E., T.R. Mempel, I.B. Mazo, B. Liu, M.N. Artyomov, H. Zheng, A. Peixoto, M.P. Flynn, B. Senman, T. Junt, et al. 2008. T cell sensing of antigen dose governs interactive behavior with dendritic cells and sets a threshold for T cell activation. *Nat. Immunol.* 9:282–291. <https://doi.org/10.1038/nri1559>
- Hilser, V.J., and E. Freire. 1996. Structure-based calculation of the equilibrium folding pathway of proteins. Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* 262:756–772. <https://doi.org/10.1006/jmbi.1996.0550>
- Hilser, V.J., and E. Freire. 1997. Predicting the equilibrium protein folding pathway: Structure-based analysis of staphylococcal nuclease. *Proteins.* 27:171–183. [https://doi.org/10.1002/\(SICI\)1097-0134\(199702\)27:2<171::AID-PROT3>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0134(199702)27:2<171::AID-PROT3>3.0.CO;2-J)
- Hilser, V.J., D. Dowdy, T.G. Oas, and E. Freire. 1998. The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble. *Proc. Natl. Acad. Sci. USA.* 95:9903–9908. <https://doi.org/10.1073/pnas.95.17.9903>
- Hilser, V.J., B. García-Moreno E, T.G. Oas, G. Kapp, and S.T. Whitten. 2006. A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* 106:1545–1558. <https://doi.org/10.1021/cr040423+>
- Inlow, D., A. Shauger, and B. Maiorella. 1989. Insect cell culture and baculovirus propagation in protein-free medium. *J. Tissue Cult. Methods.* 12: 13–16. <https://doi.org/10.1007/BF01578001>
- Jaume, J.C., S.L. Parry, A.-M. Madec, G. Sønderstrup, and S. Baekkeskov. 2002. Suppressive effect of glutamic acid decarboxylase 65-specific autoimmune B lymphocytes on processing of T cell determinants located within the antibody epitope. *J. Immunol.* 169:665–672. <https://doi.org/10.4049/jimmunol.169.2.665>
- Jenkins, M.K., and J.J. Moon. 2012. The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* 188:4135–4140. <https://doi.org/10.4049/jimmunol.1102661>
- Jurtz, V., S. Paul, M. Andreatta, P. Marcatili, B. Peters, and M. Nielsen. 2017. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199:3360–3368. <https://doi.org/10.4049/jimmunol.170089>
- Kaufmann, D.E., Bailey, P.M., Sidney, J., Wagner, B., Norris, P.J., Johnston, M.N., Cosimi, L.A., Addo, M.M., Lichterfeld, M., Altfeld, M., et al. 2004. Comprehensive Analysis of Human Immunodeficiency Virus Type 1-Specific CD4 Responses Reveals Marked Immunodominance of gag and nef and the Presence of Broadly Recognized Peptides. *J. Virol.* 78: 4463–4477. <https://doi.org/10.1128/jvi.78.9.4463-4477.2004>
- Keller, A., A. Nesvizhskii, E. Kolker, and R. Aebersold. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74:5383–5392. <https://doi.org/10.1021/ac0341261>
- Keskin, D.B., B.B. Reinhold, G.L. Zhang, A.R. Ivanov, B.L. Karger, and E.L. Reinherz. 2015. Physical detection of influenza A epitopes identifies a stealth subset on human lung epithelium evading natural CD8 immunity. *Proc. Natl. Acad. Sci. USA.* 112:2151–2156. <https://doi.org/10.1073/pnas.1423482112>
- Kim, A., and S. Sadegh-Nasseri. 2015. Determinants of immunodominance for CD4 T cells. *Curr. Opin. Immunol.* 34:9–15. <https://doi.org/10.1016/j.coi.2014.12.005>
- Kim, S.K., M. Cornberg, X.Z. Wang, H.D. Chen, L.K. Selin, and R.M. Welsh. 2005. Private specificities of CD8 T cell responses control patterns of heterologous immunity. *J. Exp. Med.* 201:523–533. <https://doi.org/10.1084/jem.20041337>
- Kim, A., I.Z. Hartman, B. Poore, T. Boronina, R.N. Cole, N. Song, M.T. Ciudad, R.R. Caspi, D. Jaraquemada, and S. Sadegh-Nasseri. 2014. Divergent paths for the selection of immunodominant epitopes from distinct antigenic sources. *Nat. Commun.* 5:5369. <https://doi.org/10.1038/ncomms6369>
- Kim, A., T.N. Boronina, R.N. Cole, E. Darrah, and S. Sadegh-Nasseri. 2017. Distorted immunodominance by linker sequences or other epitopes from a second protein antigen during antigen-processing. *Sci. Rep.* 7: 46418. <https://doi.org/10.1038/srep46418>
- Korb, L.C., S. Mirshahidi, K. Ramyar, A.A. Sadighi Akha, and S. Sadegh-Nasseri. 1999. Induction of T cell anergy by low numbers of agonist ligands. *J. Immunol.* 162:6401–6409. <https://doi.org/10.4049/jimmunol.162.11.6401>
- LaBranche, C.C., M.M. Sauter, B.S. Haggarty, P.J. Vance, J. Romano, T.K. Hart, P.J. Bugelski, M. Marsh, and J.A. Hoxie. 1995. A single amino acid change in the cytoplasmic domain of the simian immunodeficiency virus transmembrane molecule increases envelope glycoprotein expression on infected cells. *J. Virol.* 69:5217–5227. <https://doi.org/10.1128/jvi.69.9.5217-5227.1995>
- Laher, F., S. Ranasinghe, F. Porichis, N. Mewalal, K. Pretorius, N. Ismail, S. Buus, A. Stryhn, M. Carrington, B.D. Walker, et al. 2017. HIV controllers exhibit enhanced frequencies of major histocompatibility complex class II Tetramer<sup>+</sup> gag-specific CD4<sup>+</sup> T cells in chronic clade C HIV-1 infection. *J. Virol.* 91:91. <https://doi.org/10.1128/JVI.02477-16>
- Landry, S.J. 1997. Local protein instability predictive of helper T-cell epitopes. *Immunol. Today.* 18:527–532. [https://doi.org/10.1016/S0167-5699\(97\)01152-3](https://doi.org/10.1016/S0167-5699(97)01152-3)
- Lazarski, C.A., F.A. Chaves, S.A. Jenks, S. Wu, K.A. Richards, J.M. Weaver, and A.J. Sant. 2005. The kinetic stability of MHC class II:peptide complexes is a key parameter that dictates immunodominance. *Immunity.* 23: 29–40. <https://doi.org/10.1016/j.immuni.2005.05.009>
- Le, T., E.J. Wright, D.M. Smith, W. He, G. Catano, J.F. Okulicz, J.A. Young, R.A. Clark, D.D. Richman, S.J. Little, and S.K. Ahuja. 2013. Enhanced CD4<sup>+</sup> T-cell recovery with earlier HIV-1 antiretroviral therapy. *N. Engl. J. Med.* 368:218–230. <https://doi.org/10.1056/NEJMoa1110187>
- Lee, K.H., D. Xie, E. Freire, and L.M. Amzel. 1994. Estimation of changes in side chain configurational entropy in binding and folding: General methods and application to helix formation. *Proteins.* 20:68–84. <https://doi.org/10.1002/prot.340200108>
- Lindqvist, M., J. van Lunzen, D.Z. Soghoian, B.D. Kuhl, S. Ranasinghe, G. Kranias, M.D. Flanders, S. Cutler, N. Yudanin, M.I. Muller, et al. 2012. Expansion of HIV-specific T follicular helper cells in chronic HIV infection. *J. Clin. Invest.* 122:3271–3280. <https://doi.org/10.1172/JCI64314>
- Liu, J., K.L. O'Brien, D.M. Lynch, N.L. Simmons, A. La Porte, A.M. Riggs, P. Abbink, R.T. Coffey, L.E. Grandpre, M.S. Seaman, et al. 2009. Immune control of an SIV challenge by a T-cell-based vaccine in rhesus monkeys. *Nature.* 457:87–91. <https://doi.org/10.1038/nature07469>
- Liu, T., D. Pantazatos, S. Li, Y. Hamuro, V.J. Hilser, and V.L. Woods Jr. 2012. Quantitative assessment of protein structural models by comparison of H/D exchange MS data with exchange behavior accurately predicted by DXCOREX. *J. Am. Soc. Mass Spectrom.* 23:43–56. <https://doi.org/10.1007/s13361-011-0267-9>
- Lovitch, S.B., Z. Pu, and E.R. Unanue. 2006. Amino-terminal flanking residues determine the conformation of a peptide-class II MHC complex. *J. Immunol.* 176:2958–2968. <https://doi.org/10.4049/jimmunol.176.5.2958>
- Malherbe, L., C. Hausl, L. Teyton, and M.G. McHeyzer-Williams. 2004. Clonal selection of helper T cells is determined by an affinity threshold with no further skewing of TCR binding properties. *Immunity.* 21:669–679. <https://doi.org/10.1016/j.immuni.2004.09.008>
- Mempel, T.R., S.E. Henrickson, and U.H. Von Andrian. 2004. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature.* 427:154–159. <https://doi.org/10.1038/nature02238>
- Mettu, R.R., T. Charles, and S.J. Landry. 2016. CD4<sup>+</sup> T-cell epitope prediction using antigen processing constraints. *J. Immunol. Methods.* 432:72–81. <https://doi.org/10.1016/j.jim.2016.02.013>
- Milich, D.R., A. McLachlan, G.B. Thornton, and J.L. Hughes. 1987. Antibody production to the nucleocapsid and envelope of the hepatitis B virus primed by a single synthetic T cell site. *Nature.* 329:547–549. <https://doi.org/10.1038/329547a0>
- Mimura, Y., Y. Mimura-Kimura, K. Doores, D. Golgher, B.G. Davis, R.A. Dwek, P.M. Rudd, and T. Elliott. 2007. Folding of an MHC class II-restricted tumor antigen controls its antigenicity via MHC-guided processing. *Proc. Natl. Acad. Sci. USA.* 104:5983–5988. <https://doi.org/10.1073/pnas.0701307104>
- Mirano-Bascos, D., M. Tary-Lehmann, and S.J. Landry. 2008. Antigen structure influences helper T-cell epitope dominance in the human immune response to HIV envelope glycoprotein gp120. *Eur. J. Immunol.* 38:1231–1237. <https://doi.org/10.1002/eji.200738011>
- Mirshahidi, S., C.T. Huang, and S. Sadegh-Nasseri. 2001. Anergy in peripheral memory CD4<sup>+</sup> T cells induced by low avidity engagement of T cell receptor. *J. Exp. Med.* 194:719–731. <https://doi.org/10.1084/jem.194.6.719>

- Mirshahidi, S., L.C.K. Ferris, and S. Sadegh-Nasseri. 2004. The magnitude of TCR engagement is a critical predictor of T cell anergy or activation. *J. Immunol.* 172:5346–5355. <https://doi.org/10.4049/jimmunol.172.9.5346>
- Moon, J.J., H.H. Chu, M. Pepper, S.J. McSorley, S.C. Jameson, R.M.M. Kedl, and M.K. Jenkins. 2007. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity.* 27:203–213. <https://doi.org/10.1016/j.immuni.2007.07.007>
- Murphy, K.P., and E. Freire. 1992. Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* 43:313–361. [https://doi.org/10.1016/S0065-3233\(08\)60556-2](https://doi.org/10.1016/S0065-3233(08)60556-2)
- Narayan, K., C.L. Chou, A. Kim, I.Z. Hartman, S. Dalai, S. Khoruzhenko, and S. Sadegh-Nasseri. 2007. HLA-DM targets the hydrogen bond between the histidine at position  $\beta$ 81 and peptide to dissociate HLA-DR-peptide complexes. *Nat. Immunol.* 8:92–100. <https://doi.org/10.1038/nri1414>
- Narayan, K., K.W. Su, C.L. Chou, S. Khoruzhenko, and S. Sadegh-Nasseri. 2009. HLA-DM mediates peptide exchange by interacting transiently and repeatedly with HLA-DR1. *Mol. Immunol.* 46:3157–3162. <https://doi.org/10.1016/j.molimm.2009.07.001>
- Natarajan, S.K., M. Assadi, and S. Sadegh-Nasseri. 1999. Stable peptide binding to MHC class II molecule is rapid and is determined by a receptive conformation shaped by prior association with low affinity peptides. *J. Immunol.* 162:4030–4036. <https://doi.org/10.4049/jimmunol.162.7.4030>
- Ndhlovu, Z.M., S.W. Kazer, T. Nkosi, F. Ogunshola, D.M. Muema, G. Anmole, S.A. Swann, A. Moodley, K. Dong, T. Reddy, et al. 2019. Augmentation of HIV-specific T cell function by immediate treatment of hyperacute HIV-1 infection. *Sci. Transl. Med.* 11:eaa0528. <https://doi.org/10.1126/scitranslmed.aau0528>
- Nelson, C.A., S.J. Petzold, and E.R. Unanue. 1994. Peptides determine the lifespan of MHC class II molecules in the antigen-presenting cell. *Nature.* 371:250–252. <https://doi.org/10.1038/371250a0>
- Nelson, C.A., I. Vidavsky, N.J. Viner, M.L. Gross, and E.R. Unanue. 1997. Amino-terminal trimming of peptides for presentation on major histocompatibility complex class II molecules. *Proc. Natl. Acad. Sci. USA.* 94:628–633. <https://doi.org/10.1073/pnas.94.2.628>
- Nesvizhskii, A.I., A. Keller, E. Kolker, and R. Aebersold. 2003. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75:4646–4658. <https://doi.org/10.1021/ac0341261>
- Newman, J.L., E.W. Butcher, D.T. Patel, Y. Mikhaylenko, and M.F. Summers. 2004. Flexibility in the P2 domain of the HIV-1 Gag polyprotein. *Protein Sci.* 13:2101–2107. <https://doi.org/10.1110/ps.04614804>
- Niessl, J., A.E. Baxter, P. Mendoza, M. Jankovic, Y.Z. Cohen, A.L. Butler, C.L. Lu, M. Dubé, I. Shimeliovich, H. Gruell, et al. 2020. Combination anti-HIV-1 antibody therapy is associated with increased virus-specific T cell immunity. *Nat. Med.* 26:222–227. <https://doi.org/10.1038/s41591-019-0747-1>
- Pan, H., J.C. Lee, and V.J. Hilser. 2000. Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. USA.* 97:12020–12025. <https://doi.org/10.1073/pnas.220240297>
- Pauthner, M.G., J.P. Nkolola, C. Havenar-Daughton, B. Murrell, S.M. Reiss, R. Bastidas, J. Prevost, R. Nedellec, B. von Bredow, P. Abbink, et al. 2019. Vaccine-induced protection from homologous tier 2 SHIV challenge in nonhuman primates depends on serum-neutralizing antibody titers. *Immunity.* 50:241–252.e6. <https://doi.org/10.1016/j.immuni.2018.11.011>
- Poluektov, Y.O., A. Kim, I.Z. Hartman, and S. Sadegh-Nasseri. 2013. HLA-DO as the optimizer of epitope selection for MHC class II antigen presentation. *PLoS One.* 8:e71228. <https://doi.org/10.1371/journal.pone.0071228>
- Rabinowitz, J.D., M. Vrljic, P.M. Kasson, M.N. Liang, R. Busch, J.J. Boniface, M.M. Davis, and H.M. McConnell. 1998. Formation of a highly peptide-receptive state of class II MHC. *Immunity.* 9:699–709. [https://doi.org/10.1016/S1074-7613\(00\)80667-6](https://doi.org/10.1016/S1074-7613(00)80667-6)
- Ranasinghe, S., M. Flanders, S. Cutler, D.Z. Soghoian, M. Ghebremichael, I. Davis, M. Lindqvist, F. Pereyra, B.D. Walker, D. Heckerman, and H. Streeck. 2012. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol.* 86:277–283. <https://doi.org/10.1128/JVI.05577-11>
- Ranasinghe, S., S. Cutler, I. Davis, R. Lu, D.Z. Soghoian, Y. Qi, J. Sidney, G. Kranias, M.D. Flanders, M. Lindqvist, et al. 2013. Association of HLA-DRB1-restricted CD4+ T cell responses with HIV immune control. *Nat. Med.* 19:930–933. <https://doi.org/10.1038/nm.3229>
- Ranasinghe, S., D.Z. Soghoian, M. Lindqvist, M. Ghebremichael, F. Donaghey, M. Carrington, M.S. Seaman, D.E. Kaufmann, B.D. Walker, and F. Porichis. 2015. HIV-1 antibody neutralization breadth is associated with enhanced HIV-specific CD4+ T cell responses. *J. Virol.* 90:2208–2220. <https://doi.org/10.1128/JVI.02278-15>
- Ranasinghe, S., P.A. Lamothe, D.Z. Soghoian, S.W. Kazer, M.B. Cole, A.K. Shalek, N. Yosef, R.B. Jones, F. Donaghey, C. Nwonu, et al. 2016. Antiviral CD8+ T cells restricted by human leukocyte antigen class II exist during natural HIV infection and exhibit clonal expansion. *Immunity.* 45:917–930. <https://doi.org/10.1016/j.immuni.2016.09.015>
- Rice, A.P. 2017. The HIV-1 Tat protein: Mechanism of action and target for HIV-1 cure strategies. *Curr. Pharm. Des.* 23:4098–4102. <https://doi.org/10.2174/1381612823666170704130635>
- Roederer, M., Nozzi, J.L., and Nason, M.C. 2011. SPICE: Exploration and analysis of post-cytometric complex multivariate datasets. *Cytometry A.* 79:167–174. <https://doi.org/10.1002/cyto.a.21015>
- Runnels, H.A., D.A. Weber, J.C. Moore, L.E. Westerman, and P.E. Jensen. 1997. Intact proteins can bind to class II histocompatibility molecules with high affinity. *Mol. Immunol.* 34:471–480. [https://doi.org/10.1016/S0161-5890\(97\)00043-6](https://doi.org/10.1016/S0161-5890(97)00043-6)
- Russell, S.M., and F.Y. Liew. 1979. T cells primed by influenza virion internal components can cooperate in the antibody response to haemagglutinin. *Nature.* 280:147–148. (14). <https://doi.org/10.1038/280147a0>
- Sadegh-Nasseri, S., and A. Kim. 2015. Exogenous antigens bind MHC class II first, and are processed by cathepsins later. *Mol. Immunol.* 68:81–84. <https://doi.org/10.1016/j.molimm.2015.07.018>
- Sadegh-Nasseri, S., and A. Kim. 2019. Selection of immunodominant epitopes during antigen processing is hierarchical. *Mol. Immunol.* 113:115–119. <https://doi.org/10.1016/j.molimm.2018.08.011>
- Sanders, R.W., R. Derking, A. Cupo, J.P. Julien, A. Yasmeen, N. de Val, H.J. Kim, C. Blattner, A.T. de la Peña, J. Korzun, et al. 2013. A next-generation cleaved, soluble HIV-1 Env trimer, BG505 SOSIP.664 gp140, expresses multiple epitopes for broadly neutralizing but not non-neutralizing antibodies. *PLoS Pathog.* 9:e1003618. <https://doi.org/10.1371/journal.ppat.1003618>
- Sanders, R.W., M.J. van Gils, R. Derking, D. Sok, T.J. Ketas, J.A. Burger, G. Ozorowski, A. Cupo, C. Simonich, L. Goo, et al. 2015. HIV-1 VACCINES. HIV-1 neutralizing antibodies induced by native-like envelope trimers. *Science.* 349:aac4223. <https://doi.org/10.1126/science.aac4223>
- Sette, A., Adorini, L., Colon, S.M., Buus, S., and Grey, H.M. 1989. Capacity of intact proteins to bind to MHC class II molecules. *J. Immunol.* 143:1265–1267.
- Shen, R., M. Raska, D. Bimczok, J. Novak, and P.D. Smith. 2014. HIV-1 envelope glycan moieties modulate HIV-1 transmission. *J. Virol.* 88:14258–14267. <https://doi.org/10.1128/JVI.02164-14>
- Simitsek, P.D., D.G. Campbell, A. Lanzavecchia, N. Fairweather, and C. Watts. 1995. Modulation of antigen processing by bound antibodies can boost or suppress class II major histocompatibility complex presentation of different T cell determinants. *J. Exp. Med.* 181:1957–1963. <https://doi.org/10.1084/jem.181.6.1957>
- Simonetti, F.R., H. Zhang, G.P. Soroosh, J. Duan, K. Rhodehouse, A.L. Hill, S.A. Beg, K. McCormick, H.E. Raymond, C.L. Nobles, et al. 2021. Antigen-driven clonal selection shapes the persistence of HIV-1-infected CD4+ T cells in vivo. *J. Clin. Invest.* 131:131. <https://doi.org/10.1172/JCI145254>
- Soghoian, D.Z., H. Jessen, M. Flanders, K. Sierra-Davidson, S. Cutler, T. Pertel, S. Ranasinghe, M. Lindqvist, I. Davis, K. Lane, et al. 2012. HIV-specific cytolytic CD4 T cell responses during acute HIV infection predict disease outcome. *Sci. Transl. Med.* 4:123ra25. <https://doi.org/10.1126/scitranslmed.3003165>
- Stern, L.J., J.H. Brown, T.S. Jardetzky, J.C. Gorga, R.G. Urban, J.L. Strominger, and D.C. Wiley. 1994. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature.* 368:215–221. <https://doi.org/10.1038/368215a0>
- Su, L.F., B.A. Kidd, A. Han, J.J. Kotzin, and M.M. Davis. 2013. Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity.* 38:373–383. <https://doi.org/10.1016/j.immuni.2012.10.021>
- Sun, S., P. Shah, S.T. Eshghi, W. Yang, N. Trikanad, S. Yang, L. Chen, P. Aiyetan, N. Höti, Z. Zhang, et al. 2016. Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides. *Nat. Biotechnol.* 34:84–88. <https://doi.org/10.1038/nbt.3403>
- Sun, L., A.V. Paschall, D.R. Middleton, M. Ishihara, A. Ozdilek, P.L. Wantuch, J. Aceil, J.A. Duke, C.C. LaBranche, M. Tiemeyer, and F.Y. Avci. 2020. Glycopeptide epitope facilitates HIV-1 envelope specific humoral immune responses by eliciting T cell help. *Nat. Commun.* 11:2550. <https://doi.org/10.1038/s41467-020-16319-0>

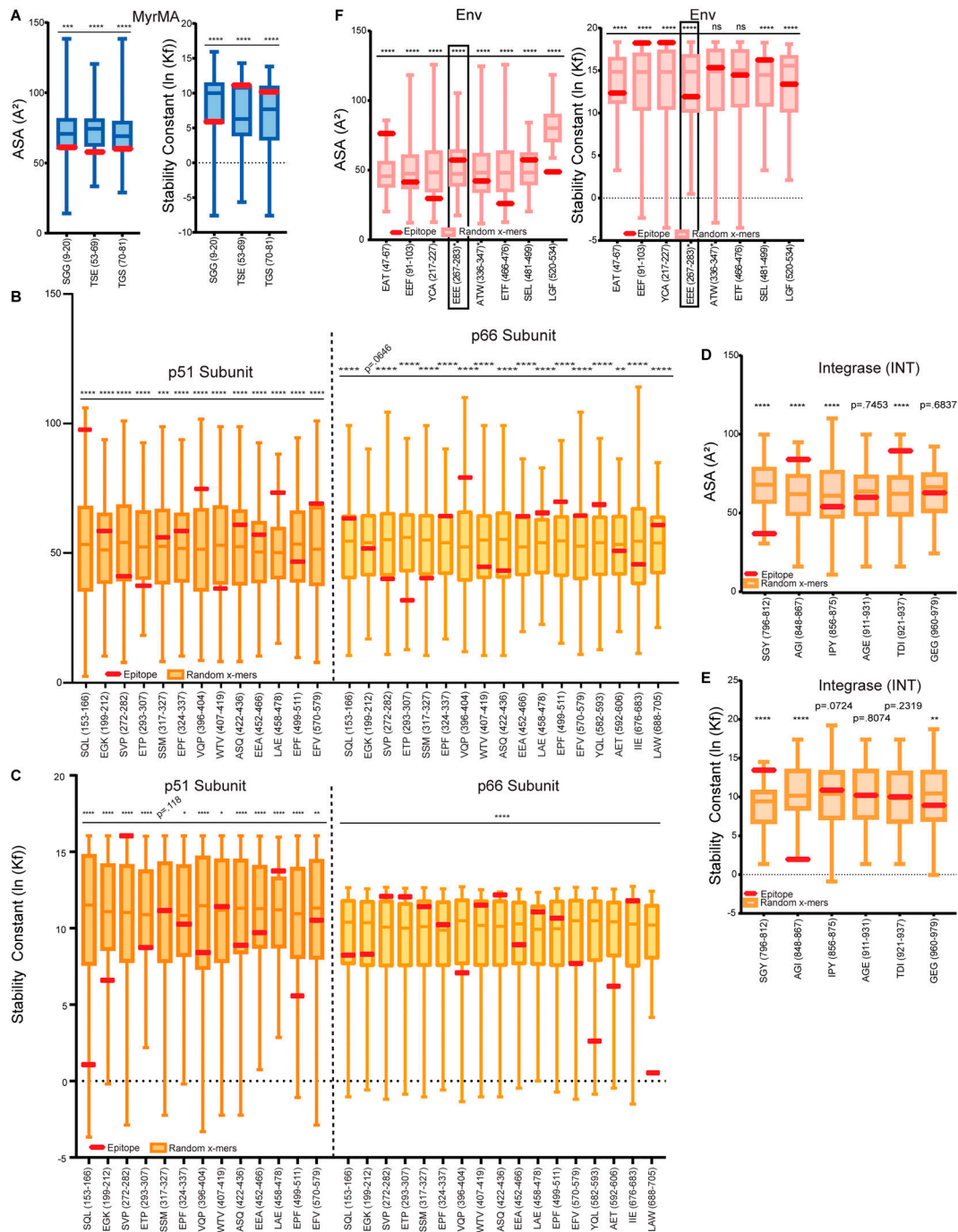


- Toghi Eshghi, S., P. Shah, W. Yang, X. Li, and H. Zhang. 2015. GPQuest: A spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact n-glycopeptides. *Anal. Chem.* 87: 5181–5188. <https://doi.org/10.1021/acs.analchem.5b00024>
- Unanue, E.R., V. Turk, and J. Neefjes. 2016. Variations in MHC class II antigen processing and presentation in health and disease. *Annu. Rev. Immunol.* 34:265–297. <https://doi.org/10.1146/annurev-immunol-041015-055420>
- Vertrees, J., P. Barritt, S. Whitten, and V.J. Hilser. 2005. COREX/BEST server: A web browser-based program that calculates regional stability variations within protein structures. *Bioinformatics.* 21:3318–3319. <https://doi.org/10.1093/bioinformatics/bti520>
- Vingert, B., S. Perez-Patrigeon, P. Jeannin, O. Lambotte, F. Boufassa, F. Le-maitre, W.W. Kwok, I. Theodorou, J.F. Delfraissy, and J. Thèze. 2010. HIV controller CD4+ T cells respond to minimal amounts of Gag antigen due to high TCR avidity. *PLoS Pathog.* 6:e1000780. <https://doi.org/10.1371/journal.ppat.1000780>
- Watts, C., and A. Lanzavecchia. 1993. Suppressive effect of antibody on processing of T cell epitopes. *J. Exp. Med.* 178:1459–1463. <https://doi.org/10.1084/jem.178.4.1459>
- Welsh, R., N. Song, and S. Sadegh-Nasseri. 2019. What to do with HLA-DO/H-2O two decades later?. *Immunogenetics.* 71:189–196. <https://doi.org/10.1007/s00251-018-01097-3>
- Welsh, R.A., N. Song, C.A. Foss, T. Boronina, R.N. Cole, and S. Sadegh-Nasseri. 2020. Lack of the MHC class II chaperone H2-O causes susceptibility to autoimmune diseases. *PLoS Biol.* 18:e3000590. <https://doi.org/10.1371/journal.pbio.3000590>
- Whitten, S.T., B. García-Moreno E, and V.J. Hilser. 2005. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci. USA.* 102:4282–4287. <https://doi.org/10.1073/pnas.0407499102>
- Whitten, S.T., A.J. Kurtz, M.S. Pometun, A.J. Wand, and V.J. Hilser. 2006. Revealing the nature of the native state ensemble through cold denaturation. *Biochemistry.* 45:10163–10174. <https://doi.org/10.1021/bi060855+>
- Wilson, C.C., B. Palmer, S. Southwood, J. Sidney, Y. Higashimoto, E. Appella, R. Chesnut, A. Sette, and B.D. Livingston. 2001. Identification and antigenicity of broadly cross-reactive and conserved human immunodeficiency virus type 1-derived helper T-lymphocyte epitopes. *J. Virol.* 75: 4195–4207. <https://doi.org/10.1128/JVI.75.9.4195-4207.2001>
- Yewdell, J.W., and J.R. Bennink. 1999. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu. Rev. Immunol.* 17:51–88. <https://doi.org/10.1146/annurev.immunol.17.1.51>
- Yin, L., J.M. Calvo-Calle, O. Dominguez-Amoroch, and L.J. Stern. 2012. HLA-DM constrains epitope selection in the human CD4 T cell response to vaccinia virus by favoring the presentation of peptides with longer HLA-DM-mediated half-lives. *J. Immunol.* 189:3983–3994. <https://doi.org/10.4049/jimmunol.1200626>
- Zybailov, B., M.K. Coleman, L. Florens, and M.P. Washburn. 2005. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* 77:6218–6224. <https://doi.org/10.1021/ac050846r>

## Supplemental material







**Figure S2. Accessibility and stability trends of HIV-1 epitopes from cell-free antigen processing.** (A) Epitopes from Myr-MA were analyzed by PDB PISA and COREX/BEST algorithms. Box-and-whiskers plots show distribution of average solvent ASA (left) or stability constants (right) of random x-mers spanning the protein (derived via sliding scale analysis, see Materials and Methods). These distributions exclude the EOI, shown in red so that its accessibility/stability relationship to the rest of the protein can be visualized. In these plots, higher ASA values indicate higher epitope exposure and lower lnKf values indicated decreased epitope stability. (B and C) Box-and-whiskers plots showing the distribution of (B) average ASA or (C) average stability constants from randomly generated p51 and p66 epitopes excluding the EOI. EOI in red. Stability constants were determined using COREX based on the 1HMY structure of RT, analyzing each subunit independently. (D and E) Box-and-whiskers plots showing the distribution of average (D) ASA or (E) stability of random x-mer epitopes spanning the INT protein (1EX4) and excluding the EOI. (F) Box-and-whiskers plot showing the distribution of average ASA (PDB PISA; left) or stability (COREX; right) of random x-mer epitopes spanning the BG505 SOSIP protein (4ZMJ) and excluding the EOI, with a box highlighting the EEE glycopeptide. EOI is shown in red. Normally distributed data in A–F were subject to a one-sample, two-tailed *t* test, and non-normally distributed data were subject to a two-tailed Wilcoxon Signed Rank Test, comparing the mean (*t* test) or median (Wilcoxon Signed Rank) of the distribution to the mean ASA or stability of the epitope. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001; \*\*\*\**P* < or = 0.0001.

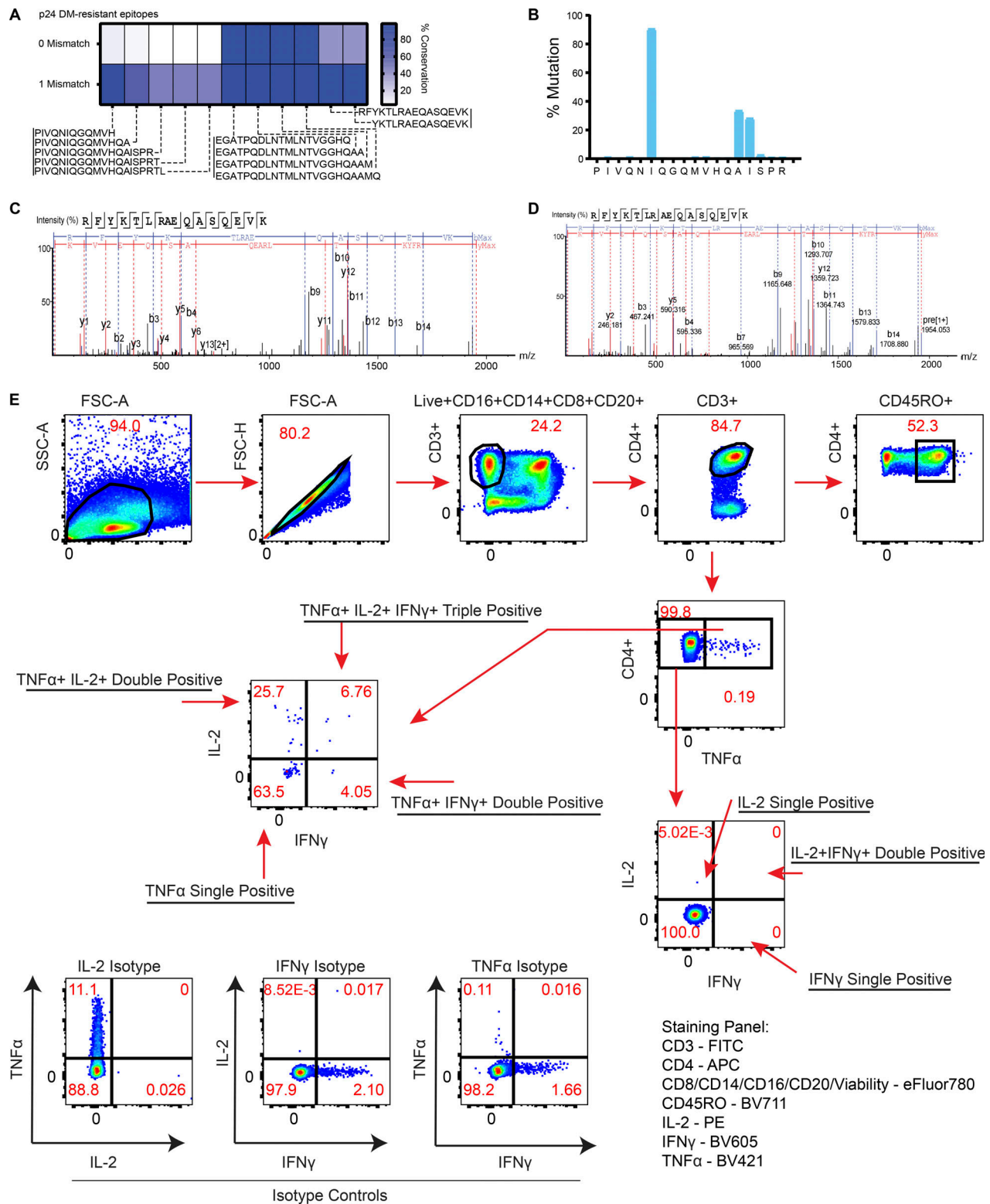
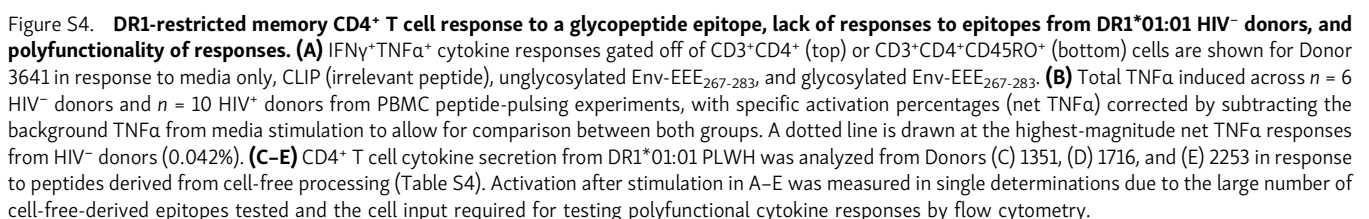
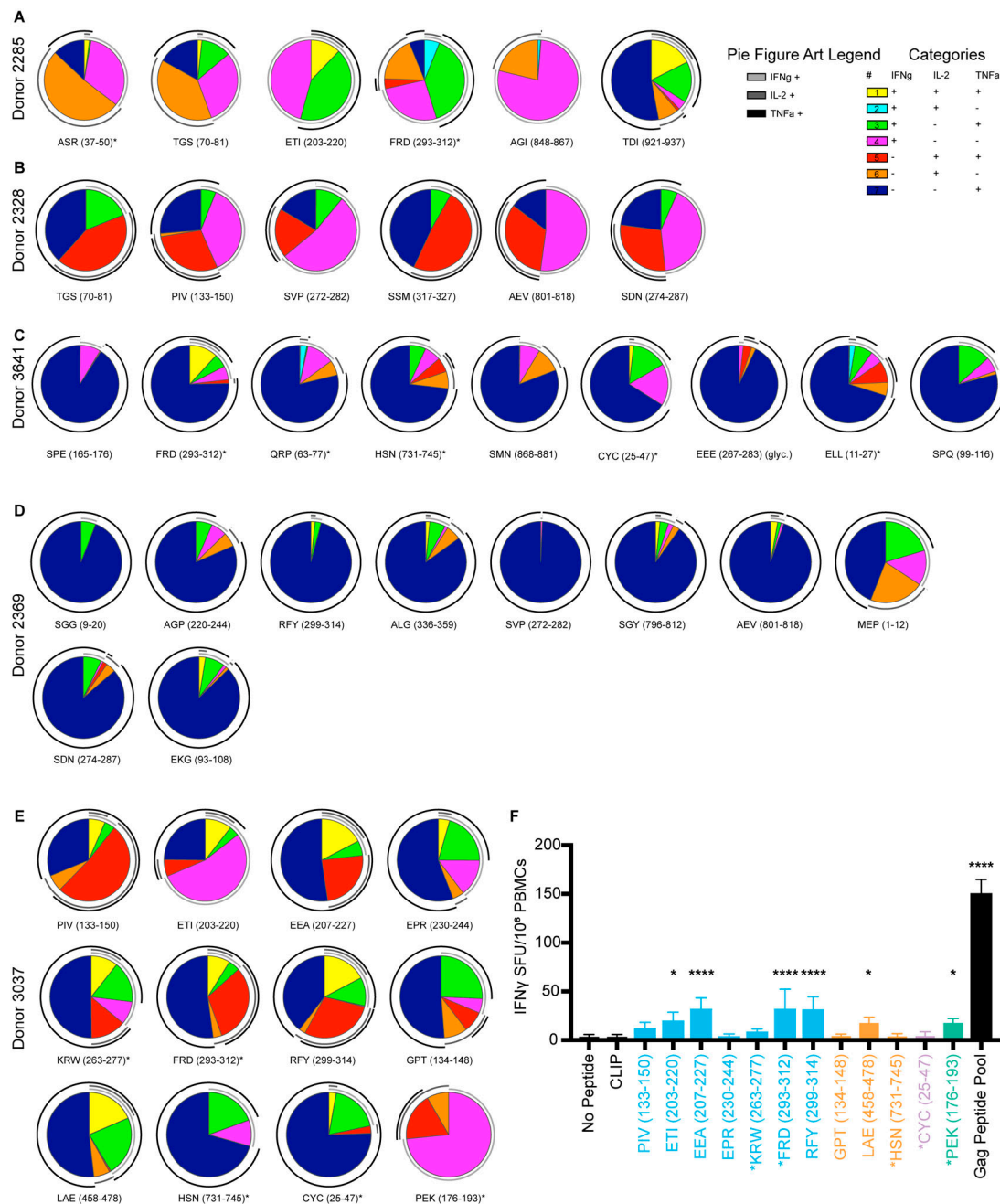


Figure S3. **Mutational and peptide detection characteristics of cell-free-derived HIV-1 epitopes, and gating scheme for measuring CD4<sup>+</sup> T cell responses from PLWH.** (A) Heat map showing epitope sequence prevalence of all epitopes obtained from cell-free processing of p24 compared with 1066 Clade B sequences (LANL), with either 0 or 1 aa mismatch tolerated. (B) Percentage of mismatches per residue is shown for Gag-PIV<sub>133-150</sub> from p24-HXB2. (C) MS/MS scan shows b/y ions for a low-abundance, DM-resistant epitope RFYKTLRAEQASQEVK (Gag-RFY<sub>299-314</sub>) from p24 cell-free processing identified with 1 PSM. (D) MS/MS scan of same epitope (Gag-RFY<sub>299-314</sub>), this time synthesized and spiked in at 25 femtomole quantity into murine B cell peptide elution samples containing self-peptides. (E) Representative gating strategy for single-, double-, and triple-positive intracellular cytokine responses gated off of Live<sup>+</sup>CD3<sup>+</sup>CD4<sup>+</sup> T after 16 h of PBMC incubation with single HIV-1 peptides from cell-free processing. Fluorescence minus one controls containing the relevant isotype antibodies were used for gating for cytokines, and CD45RO<sup>+</sup> was used to confirm that responses were in the memory compartment.







**Figure S5. Polyfunctionality of CD4<sup>+</sup> T cell responses in PLWH to peptides derived from cell-free processing and IFN $\gamma$  ELISPOT validation for Donor 3037. (A–E)** CD4<sup>+</sup> T cell cytokine secretion from DR1\*01:01 PLWH was analyzed from Donors (A) 2285, (B) 2328, (C) 3641, (D) 2369, and (E) 3037 in response to peptides derived from cell-free processing. Peptides are listed as in Table S4, with asterisks denoting control peptides identified in the literature. Double-positive cells were those that produced two cytokines. Double positive cells were calculated by the following strategies: (a) IL-2<sup>+</sup>TNF $\alpha$ <sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ TNF $\alpha$ <sup>+</sup>→ IL-2<sup>+</sup>, and the percentage of IL-2<sup>+</sup> cells was multiplied by the percentage of TNF $\alpha$ <sup>+</sup> cells from the parent gate; (b) IFN $\gamma$ <sup>+</sup>TNF $\alpha$ <sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ TNF $\alpha$ <sup>+</sup>→ IFN $\gamma$ <sup>+</sup>, and the percentage of IFN $\gamma$ <sup>+</sup> cells were multiplied by the percentage of TNF $\alpha$ <sup>+</sup> cells from the parent gate; (c) IL-2<sup>+</sup>IFN $\gamma$ <sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ IL-2<sup>+</sup>IFN $\gamma$ <sup>+</sup>→ TNF $\alpha$ <sup>+</sup>, and the percentage of TNF $\alpha$ <sup>+</sup> cells were multiplied by the percentage of IL-2<sup>+</sup>IFN $\gamma$ <sup>+</sup> cells from the parent gate. Single-positive populations were calculated by the following: (a) IL-2<sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ TNF $\alpha$ <sup>-</sup>→ IL-2<sup>+</sup>IFN $\gamma$ <sup>-</sup>, and the percentage of IL-2<sup>+</sup>IFN $\gamma$ <sup>-</sup> cells were multiplied by the percentage of TNF $\alpha$ <sup>-</sup> cells from the parent gate; (b) IFN $\gamma$ <sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ TNF $\alpha$ <sup>-</sup>→ IL-2<sup>-</sup>IFN $\gamma$ <sup>+</sup>, and the percentage of IL-2<sup>-</sup>IFN $\gamma$ <sup>+</sup> cells were multiplied by the percentage of TNF $\alpha$ <sup>-</sup> cells from the parent gate; (c) TNF $\alpha$ <sup>+</sup>: Lymphocytes→ Single Cells→ Live<sup>+</sup>CD3<sup>+</sup>→ CD3<sup>+</sup>CD4<sup>+</sup>→ TNF $\alpha$ <sup>+</sup>→ IL-2<sup>-</sup>IFN $\gamma$ <sup>-</sup>, and the percentage of IL-2<sup>-</sup>IFN $\gamma$ <sup>-</sup> cells were multiplied by the percentage of TNF $\alpha$ <sup>+</sup> cells from the parent gate. Data gated in this manner was analyzed by SPICE and represented as pie charts in A–E. Responses depicted are to those peptides that induced CD4<sup>+</sup> T cell cytokine responses that were 2.95-fold greater than the response to CLIP. Activation after stimulation in A–E was measured in single determinations. **(F)** IFN $\gamma$  ELISPOT from Donor 3037 with data shown as spot forming units (SFU) per million PBMCs. Peptides utilized were those shown in E. Responses to negative controls (no peptide, CLIP) and positive control (Gag peptide pool) are shown as comparators, with six replicate wells run for each condition. Data represents two independent experiments. Significance difference relative to CLIP determined by one-way ANOVA with Dunnett's test for multiple comparisons, \*P < 0.05, \*\*\*\*P < 0.0001.

Provided online are nine tables. Table S1 shows HIV proteins subjected to cell-free processing used in this study. Table S2 shows epitopes obtained from cell-free processing reported in previous studies used in this study. Table S3 shows glycopeptides obtained from cell-free processing identified by GPQuest used in this study. Table S4 lists select peptides identified from cell-free processing used in this study. Table S5 describes the clinical characteristics of PLWH in the study. Table S6 lists sequences of proteins utilized in the cell-free processing system used on this study. Table S7 provides raw data obtained from cell-free processing (Excel) used in this study. Table S8 provides an example of COREX analysis for Myr-MA (Excel) used in this study. Table S9 provides the CD4<sup>+</sup> T cell responses from PLWH (shown as fold change over CLIP) in ex vivo stimulations (Excel) used in this study.