The nature of autoantigens that trigger autoimmune diseases has been much discussed, but direct biochemical identification is lacking for most. Addressing this question demands unbiased examination of the self-peptides displayed by a defined autoimmune major histocompatibility complex class II (MHC-II) molecule. Here, we examined the immunopeptidome of the pancreatic islets in non-obese diabetic mice, which spontaneously develop autoimmune diabetes based on the I-Aβ variant of MHC-II. The relevant peptides that induced pathogenic CD4+ T cells at the initiation of diabetes derived from proinsulin. These peptides were also found in the MHC-II peptidome of the pancreatic lymph nodes and spleen. The proinsulin-derived peptides followed a trajectory from their generation and exocytosis in β cells to uptake and presentation in islets and peripheral sites. Such a pathway generated conventional epitopes but also resulted in the presentation of post-translationally modified peptides, including deamidated sequences. These analyses reveal the key features of a restricted component in the self-MHC-II peptidome that caused autoreactivity.
was distinctly different from that of I-A\(^d\) NOD mice (Fig. 1a), indicating the specificity of the I-A\(^d\) peptidome.

We isolated ~4 \times 10^6 APCs from the islets of the I-A\(^d\) NOD mice and detected MHC-II-bound peptides from 112 proteins. The islet MHC-II peptidome was enriched in \(\beta\) cell-derived peptides identified in the \(\beta\) cell secretory granules (Table 1 and Supplementary Table 1), mostly derived from \(\beta\) cell secretory granules (Table 1 and Supplementary Table 1). The \(\beta\) cell-derived peptides were grouped in families, each having sequences of different lengths and variable flanking residues (Table 1 and Supplementary Table 1). Most of the \(\beta\) cell-derived peptides were found in islets, with 20 residues or more (Extended Data Fig. 1c).

InsB and InsC dominated the islet MHC-II peptidome in relative abundance and in the large number of peptide variants (Table 1). Intact InsB (InsB:1–30; 31 residues) and InsC of proinsulin-1 (Ins1C:33–61; 29 residues) were detected among the MHC-II-bound epitopes (Supplementary Table 1), raising the question of whether these were further processed by the APCs to smaller fragments or bound as such to MHC-II. Using a standard antigen presentation assay that measures interleukin-2 (IL-2) production by CD4\(^+\) T cell hybridomas during antigen-specific stimulation \(^{17,21,25}\), we found that treatment of the C3g7 APCs (a B cell lymphoma cell line expressing I-A\(^d\)) with chloroquine (a drug that prevents intracellular processing) did not affect the presentation of Ins1C:33–61 to a cognate CD4\(^+\) T cell hybridoma (Extended Data Fig. 1d). Thus, the long MHC-II-bound peptides can be directly presented without a reduction in their length. A hybrid insulin peptide (HIP) formed by the fusion of InsC with the islet amyloid polypeptide (IAPP), which had a sequence (LQTLAL–NAARD) identical to that of peptide HIP6.9 (refs. \(^{26,27}\)), was also identified in the MHC-II peptidome (Table 1). This InsC–IAPP HIP was detected at a low level in the islet MHC-II peptidome and represented <1% of the total peptide in each case (Table 1 and Supplementary Tables 1 and 2).

Considering the important role of peripheral presentation in the development of autoimmune diabetes \(^{21,26,28}\), we examined the MHC-II peptidome of the pLN (Supplementary Table 2) and spleen (Supplementary Table 3), as well as the free peptides in DCGs and crinosomes and in the secreted peptides from \(\beta\) cells in response to glucose challenge \(^{21}\) (referred to as the secretome hereafter; Supplementary Table 4). Several \(\beta\) cell peptide families were identified in the MHC-II peptidome of the pLN and spleen, mostly from

---

**Fig. 1** Analysis of the \(\beta\) cell-derived peptides identified in the MHC-II peptidome. **a**, Epitope mapping by Gibbs cluster analysis of all of the MHC-II-bound peptides from the spleens of 8- to 10-week-old NOD.I-A\(^d\) (left) and NOD.I-A\(^d\) mice (right). **b–d**, Individual \(\beta\) cell-derived peptides identified in the MHC-II peptidome of the pancreatic islets and peripheral lymphoid tissues (islets + periphery; left) or in the islets alone (islets only; right) were examined for their corresponding T cell autoreactivity (\(b\)), relative MHC-II-binding affinity (\(c\)) and relative abundance (\(d\)). For **b**, an ELISPOT assay was performed to measure IL-2 and IFN-\(\gamma\) production by CD4\(^+\) T cells from islets and pLNs in 8- to 10-week-old female NOD mice upon challenging and recalling with the indicated peptides. NE, not examined. The data in **b** (means ± s.e.m.) are from \(n=5\) mice per independent experiment (each data point). For **c**, a competitive binding assay was performed to detect the relative MHC-II-binding affinity of the indicated peptides. The data in **c** (means) are from \(n=2\) independent experiments with similar results. For **d**, the abundance (normalized peak area) of the indicated peptides relative to the most abundant peptide (Ins2C:33–63) was measured. Data in **d** (means) are from the mass spectrometry analysis of the MHC-II peptidome of the pancreatic islets isolated from \(n=219\) female NOD mice of 8-10 weeks of age. **e**, Results of an ELISPOT assay, showing IL-2 and IFN-\(\gamma\) production by CD4\(^+\) T cells from islets and pLNs in female NOD mice of different ages upon challenging and recalling with the indicated peptides. Data (means ± s.e.m.) are from \(n=5\) mice per independent experiment (each data point).
InsB and InsC; others included IAPP, zinc transporter 8 (ZnT8) and the hybrid InsC–IAPP, in lesser amounts (Table 1). We also examined the MHC-II peptidome of CD19<sup>+</sup> B cells isolated from the spleens of 8- to 10-week-old female NOD mice as well as from the spleens of 8- to 10-week-old female B cell-deficient (µMT) NOD mice. The peptide families are ranked according to their relative abundance (the cumulative peak area of all of the peptides in a given family normalized to the most abundant family) in the islet MHC-II peptidome. The number of individual peptide sequences belonging to each family is shown. Data are from n = 219 mice from one experiment (islets); n = 257 mice from two independent experiments (pLN); n = 16 mice from four independent experiments (spleen); n = 4 mice from one experiment (B cells); and n = 4 mice from one experiment (µMT). The free peptides are identified in the DCGs and crinosomes, as well as in β cell secretion upon glucose challenge (secretome). Data are from reanalysis of previously published datasets (ref. 21).

### Table 1 | β cell-derived peptide families identified in the MHC-II peptidomes

| Description | Protein coverage | Relative abundance (islets) | MHC-II-bound peptides | Free peptides |
|-------------|-----------------|----------------------------|-----------------------|--------------|
|             | Protein coverage | Whole | Islets | pLN | B cells | µMT | Spleen | DCG | Crinosome | Secretome |
| Insulin-2 C-peptide | Ins2C:33–63 | 1.00 | 16 | 38 | 23 | 31 | 41 | 129 | 160 | 227 |
| Insulin-1 C-peptide | Ins1C:33–61 | 0.82 | 13 | 30 | 16 | 7 | 20 | 108 | 133 | 182 |
| Neuroendocrine protein 7B2 | Sgc5:27–57 | 0.55 | 8 | 4 | 13 | 31 |
| Islet amyloid polypeptide | IAPP:78–93 | 0.120 | 13 | 3 | 7 | 23 | 32 | 36 |
| Insulin-2 B-chain | Ins2B:1–23 | 0.030 | 7 | 4 | 3 | 5 (2) | 7 (3) | 3 | 14 (13) | 28 (22) |
| Chromogranin A | Chga:358–377 | 0.027 | 2 | 2 | 0 (2) | 3 (0) | 5 (1) | 2 | 6 (5) | 18 (12) |
| Insulin-1 B-chain | Ins1B:1–30 | 0.017 | 2 | 2 | 0 (0) | 3 (0) | 5 (1) | 2 | 6 (5) | 18 (12) |
| Zinc transporter B | ZnT8:313–322 | 0.011 | 2 | 1 | 1 | 1 |
| Chromogranin A | Chga:393–402 | 0.006 | 2 | 6 | 11 | 12 |
| Chromogranin A | Chga:19–29 | 0.005 | 2 | 6 | 11 | 12 |
| Vitamin D-binding protein | DBP:389–405 | 0.005 | 1 | 11 | 5 |
| Hybrid InsC–IAPP | InsC–IAPP | 0.004 | 1 | 1 |
| Secretogranin-1 | Scg1:64–86 | 0.004 | 1 | 1 | 4 | 11 |
| Secretogranin-1 | Scg1:516–535 | 0.003 | 1 | 1 | 11 | 16 |
| Secretogranin-1 | Scg1:333–355 | 0.003 | 1 | 1 |
| Secretogranin-2 | Scg2:28–41 | 0.002 | 1 | 1 | 9 | 9 |
| Secretogranin-2 | Scg2:184–216 | 0.001 | 1 | 7 | 19 | 40 |
| 60-kD heat-shock protein | HSP:167–181 | 0.001 | 2 | 12 | 13 | 10 | 2 |
| Secretogranin-2 | Scg2:569–610 | <0.001 | 1 | 10 | 28 | 57 |

<sup>a</sup>One or more peptides did not pass the 1% FDR cutoff but were manually verified. <sup>b</sup>One or more peptides contained a C19S substitution. Values in parentheses show the number of unique sequences specifically belonging to the insulin-1 or insulin-2 B-chain family. The β cell-derived peptide families are identified in the MHC-II peptidome of the islets, pLNs, spleens and CD19<sup>+</sup> splenic B cells isolated from 8- to 10-week-old female NOD mice, as well as from the spleens of 8- to 10-week-old female B cell-deficient (µMT) NOD mice. The peptide families are ranked according to their relative abundance (the cumulative peak area of all of the peptides in a given family normalized to the most abundant family) in the islet MHC-II peptidome. The number of individual peptide sequences belonging to each family is shown. Data are from n = 219 mice from one experiment (islets); n = 257 mice from two independent experiments (pLN); n = 16 mice from four independent experiments (spleen); n = 4 mice from one experiment (B cells); and n = 4 mice from one experiment (µMT). The free peptides are identified in the DCGs and crinosomes, as well as in β cell secretion upon glucose challenge (secretome). Data are from reanalysis of previously published datasets (ref. 21).

Selected β cell-derived peptides induce autoreactivity. We carried out a screen of the representative peptides from each β cell-derived peptide family (Table 2) for their corresponding CD4<sup>+</sup> T cell reactivity and relative MHC-II-binding strength. Each synthetic β cell-derived peptide (Table 2) was added for two challenge cycles to pooled cells isolated from the islets and pLNs of 8- to 10-week-old female NOD mice. The reactivity of the CD4<sup>+</sup> T cells was assayed by measuring the production of the cytokines IL-2 and interferon-γ (IFN-γ) by enzyme-linked immune absorbent spot (ELISPOT). We also measured the production of IL-10 and IFN-4, which were invariably negative (results not shown). Stimulation of the islet and pLN CD4<sup>+</sup> T cells with an I-Aβ<sup>+</sup>-binding hen egg-white lysozyme (HEL) peptide did not detect CD4<sup>+</sup> T cells specific for this foreign antigen (Extended Data Fig. 2a), indicating that the assay was specific and did not generate de novo responses. We also examined the sensitivity of the assay by spiking in pre-activated HEL-reactive CD4<sup>+</sup> T cells and detected as few as 20 CD4<sup>+</sup> T cells (Extended Data Fig. 2b), indicating the high sensitivity of the assay. The relative binding strength of the peptides to MHC-II was evaluated in a competitive binding assay against the reference peptide (g7-MIME) known to strongly bind to I-Aβ<sup>+</sup> (Extended Data Fig. 2c).

Only a limited number (five out of 21 examined) of β cell-derived peptides correlated with the presence of autoreactive CD4<sup>+</sup> T cells in islets and pLNs (Fig. 1b). These MHC-II-bound peptides were identified in both islets and the peripheral lymphoid tissues (Fig. 1b and Table 2). The major populations of reactive CD4<sup>+</sup> T cells recognized the peptide 9–23 of InsB (InsB:9–23) and several Ins1C segments, including Ins1C:33–61, Ins1C:37–61 and Ins1C:51–61 (Fig. 1b), while the intact InsC of proinsulin-2 (Ins2C:33–63) induced little response (Fig. 1b). We also detected CD4<sup>+</sup> T cells specific to the hybrid peptide InsC–IAPP (Fig. 1b). The peptides identified only in...
We identified variants of InsB-derived peptides, each having divergent biological properties \(^{17,25,32}\), with specificities toward the islets, pLNs, and spleen (Fig. 2a). The InsB:9–23-reactive CD4\(^+\) T cells in the islets and pLNs of 8- to 10-week-old female NOD mice responded to InsB:12–20, but not InsB:13–21 (Fig. 2b). No T cell reactivity was detected for two other putative registers, InsB:14–22 and InsB:15–23 (Fig. 2b), indicating that the responses were highly specific to InsB:12–20.

We also identified three other InsB peptides containing the InsB:12–20 register; namely, InsB:11–25, InsB:12–25 and InsB:12–26, in the spleen MHC-II peptidome. The MHC-II epitope predictive program TEA-DQ8 \(^{33}\) suggested that InsB:15–23 was the preferred binding register included in InsB:11–25, followed by InsB:12–20 (Extended Data Fig. 3a). We generated 46 CD4\(^+\) T cell hybridomas reactive to InsB:11–25 and found that half of these (represented by clone 21) cross-reacted with InsB:12–20 (Fig. 2c). The rest (represented by clone 58) did not recognize either InsB:12–20 or InsB:13–21 (Fig. 2c) but responded to InsB:15–23 (Extended Data Fig. 3b), although the recognition was highly dependent on the amino- and carboxy-flanking residues (Extended Data Fig. 3c). However, when the islet and pLN CD4\(^+\) T cells were challenged with InsB:11–25, only CD4\(^+\) T cells reactive to InsB:12–20 were identified (Fig. 2d). The islet APCs isolated from the same mice spontaneously activated clone 21, but not clone 58 (Fig. 2e). Thus, InsB:15–23-specific CD4\(^+\) T cells did not directly contribute to the autoimmune process.

InsB:11–25 bound to I-A\(^\*\) \(^{34}\) stronger than InsB:9–23 (Fig. 2f). A brief pulse of the C3g7 APCs with InsB:11–25 resulted in sustained presentation of InsB:12–20 to the 9B9 CD4\(^+\) T cell hybridoma (Fig. 2g), which has specificity to InsB:12–20 \(^{35}\). In contrast, the responses to InsB:9–23 declined rapidly (Fig. 2g). These results indicated that InsB:11–25 was a better carrier of the InsB:12–20 epitope. Primary CD4\(^+\) T cells activated with InsB:11–25, which contained InsB:12–20-reactive T cells, transferred diabetes into NOD Rag1\(^{-/-}\) recipient mice (Fig. 2h). In summary, the MHC-II peptidome contains a family of InsB peptides but the pathogenic CD4\(^+\) T cells center on recognition of the InsB:12–20 epitope.

### Table 2 | Representative \(\beta\) cell-derived peptides identified in the MHC-II peptidomes

| Sites + periphery | Peptide | Sequence |
|-------------------|---------|----------|
| InsB:9–23         | SMLVEALVLCGERG | |
| InsC:33–61        | EVEDQVEQLEGSGPGDLQTALAEVARQ | |
| InsC:37–56        | PQVQLEGSGPGDLQTALAEVARQ | |
| InsC:35–61        | PQVQLEGSGPGDLQTALAEVARQ | |
| InsC:51–61        | LQTLAEVARQ | |
| InsC:53–61        | TLAEVARQ | |
| InsC:23–63        | EVEDQVAALEGGPGGGDLQTALAEVARQ | |
| IAPP:78–90        | NAARDPRNRLDF | |
| InsC-IAPP         | LQTLANNAARD | |
| ZnT8:313–326      | ILSVHVTAAASQD | |
| DBP:389–405       | SPLLRKQTSFIEKQEE | |
| HSP:367–181       | EIAQVATISANDGKD | |

| Sites only         | Peptide | Sequence |
|-------------------|---------|----------|
| ChgA:358–371      | WSRMDQALKALETKR | |
| Sgc1:3A–86        | SGKEVGKEEGQGNSKFEVRL | |
| Sgc3:333–355      | SHHLAHYRASEEPEYGEESRY | |
| Sgc1:438–454      | LLDQHYPVRESIPDTA | |
| Sgc1:516–535      | LGALNPYDPQDQKWSDFE | |
| Sgc2:228–41       | ASFQRQNLQKEDP | |
| Sgc2:287–297      | SGQQLPDEEN | |
| Sgc2:569–578      | IPVGLSFLNED | |
| Sgc5:27–57        | YSPRTPDRVSETDQIRLHGVMEOQLIARPR | |

Shown are representative peptide sequences from each peptide family depicted in Table 1. The peptides are identified in the MHC-II peptidome of both islets and peripheral lymphoid tissues (islet + periphery) or only in the islets (islets only).

Diverse InsB peptides harbor a pathogenic epitope. CD4\(^+\) T cells reactive to InsB:9–23 \(^{13,15}\) and two distinct nine-amino-acid register included in InsB:9–23 (specifically, InsB:12–20 and InsB:13–21), each having divergent biological properties \(^{17,25,32}\), have been previously described. The InsB:12–20-specific CD4\(^+\) T cells are reactive only to extracellularly derived InsB peptides bound to I-A\(^\*\) \(^{13,15}\) and are highly diabetogenic \(^{17,25,32}\). The InsB:13–21-reactive CD4\(^+\) T cells recognize APCs that process native insulin, are of a low frequency and are largely unresponsive upon antigen challenge \(^{17,25,32}\). We identified variants of InsB-derived peptides centered on the 9–23 region, including InsB:9–23, across the islets, pLN and spleen (Fig. 2a). The InsB:9–23-reactive CD4\(^+\) T cells in the islets and pLNs of 8- to 10-week-old female NOD mice responded to InsB:12–20, but not InsB:13–21 (Fig. 2b). No T cell reactivity was detected for two other putative registers, InsB:14–22 and InsB:15–23 (Fig. 2b), indicating that the responses were highly specific to InsB:12–20.

The islets generated low or negative CD4\(^+\) T cell responses (Fig. 1b and Table 2), showed relatively low MHC-II-binding strength (Fig. 1c) and were in low abundance (Fig. 1d). CD4\(^+\) T cells reactive to peptides from three known autoantigens, including chromogranin A (ChgA:358–371), ZnT8:313–326 and IAPP:78–90 were detected only in the islets and pLNs from 16- and 20-week-old female NOD mice (Fig. 1e). Because these peptides were already found in the MHC-II peptidome at 8–10 weeks (Table 2), 'epitope spreading' may result from the appearance of de novo pMHCs, but could be explained by the particular biology of the T cells, such as their diversification following initial inflammation. CD4\(^+\) T cells specific for peptides from secretogranin-1 (Sgc1:438–454 and Sgc1:516–535) and vitamin D-binding protein (DBP:389–405) were not detected throughout the time course (Fig. 1e). In brief, most of the CD4\(^+\) T cell autoreactivity in islets and pLNs at the initial stage of diabetes is directed to peptides derived from InsB and InsC.

### InsC peptides give rise to autoreactive CD4\(^+\) T cells.

Next, we examined the CD4\(^+\) T cell responses directed to InsC1 peptide segments encompassing their C-terminal region. Most of the MHC-II-bound Ins1C- and Ins2C-derived peptides spanned the amino-terminal region, while only a few contained the C terminus (Fig. 3a). Production of IFN-\(\gamma\) and IL-2 by CD4\(^+\) T cells from the draining lymph nodes 7 d after immunization of NOD mice with the long peptide Ins1C:33–63 was only detected after restimulation with segments containing the C-terminal segment (Ins1C:33–63 and Ins1C:49–63), but not with Ins2C:33–65 (Fig. 3b). Immunization with Ins1C:41–55 or Ins1C:37–56, which lacks the C terminus, did not generate any response (Extended Data Fig. 4a). Adoptive transfer of primary CD4\(^+\) T cells specific for Ins1C:49–63 into NOD Rag1\(^{-/-}\) mice induced diabetes in the recipient mice (Fig. 3c), indicating that presentation of the Ins1C C-terminal peptides induced pathogenic T cell responses. We identified deamidation of the four glutamine residues in the Ins1C peptides in all of the MHC-II peptidomes, including the glutamine to glutamic acid change at the C terminus (Fig. 3a and Supplementary Tables 1–3). Deamidation in the native Ins1C:51–61 (LQTLAEVARQ to LQTLAEVARQ) was found in the spleen, while deamidation was only detected in Ins1C:53–61E (TLAEVARQ to LQTLAEVARQ) in islets (Fig. 3a and Supplementary Tables 1–3). Deamidation in the C terminus of Ins2C:53–63 (LQTLAEVARQ) was not found across the three sites (Fig. 3a and Supplementary Tables 1–3). Because deamidation and citrullination (arginine to citrulline) result in the same mass change, we compared the spectrum of the synthetic deamidated Ins1C:53–61E with that of the synthetic citrullinated peptide (TLAEVARQ; r: citrulline) and found that the peptide spectrum from the islet peptide matched deamidation, but not citrullination (Fig. 3d and Extended Data Fig. 4b).

To test whether deamidation could have been introduced during the isolation of the peptide, we subjected the native Ins1C:53–61 (TLAEVARQ) to all of the procedures for isolation of the MHC-II
peptidome. The deamidated version (TLALEVARE) was not detected by mass spectrometry (result not shown). Furthermore, a BLAST search with the TLALEVARE sequence indicated matches to several microbial species found in the mouse gut microbiome (results not shown). To determine whether the deamidated Ins1C:53–61E was of bacterial origin, we examined the MHC-II peptidome of germ-free NOD mice. Ins1C:53–61E was detected in the peptidome. The deamidated version (TLALEVARE) was not found in the peptidome, indicating that it is of mammalian origin.

To understand how deamidation may impact autoreactivity, we compared the immunogenicity of the deamidated Ins1C:51–61E, which was identified only in the periphery (Supplementary Table 3), with native Ins1C:51–61. First, Ins1C:51–61E displayed a roughly threefold higher binding affinity to I-A<sup>β</sup> than Ins1C:51–61 (Fig. 3e), confirming that peptides with acidic residues are preferentially bound by I-A<sup>β</sup> (ref. 9). The short Ins1C:53–61E bound poorly to I-A<sup>β</sup> (Fig. 3e), indicating the importance of flanking residues. Next, we generated a panel of CD4<sup>+</sup> T cell hybridomas reactive to Ins1C:51–61E, and found that all of them cross-reacted to the native Ins1C:51–61 (Fig. 3f); however, significantly lower amounts of Ins1C:51–61E were required to reach the half-maximal T cell responses (Fig. 3f). Finally, we searched the islets and pLNs of 8- to 10-week-old female NOD mice for CD4<sup>+</sup> T cells specific for Ins1C:51–61 or Ins1C:51–61E. Ins1C:51–61E generated substantially stronger responses than native Ins1C:51–61 or Ins1C:33–61 (Fig. 3g). In summary, deamidation changes in InsC did not give rise to CD4<sup>+</sup> T cells that exclusively recognized this post-translational modification (PTM), but they enhanced the immunogenicity of the unmodified peptide.

A multi-round search identifies an MHC-II-bound HIP. Because hybrid peptides are not included in standard canonical protein databases, we employed a multi-round search strategy. We first searched a standard UniProt/SwissProt mouse proteome database and then a published database of in silico-generated HIPs<sup>34</sup>. If a spectrum was identified as a possible match to a HIP sequence, we then checked for possible alternative assignments by querying PEAKS PTM<sup>35</sup> and SPIDER<sup>36</sup> to determine whether other matches for a given spectrum were present. This strategy required the consideration of alternative sequences, such as those generated from non-standard modifications of the sample. Without such considerations, unknown sequences could be falsely assigned.

Five potential HIPS from the MHC-II-bound peptides isolated from islets and pLNs were searched for alternative assignments (Table 3). The spectra for four of these peptides were better assigned to native or modified InsC fragments (Table 3), indicating that the initial assignment as HIP was incorrect. In addition, the spectra of these four peptides did not match those of the synthetic HIP standards (Table 3 and Extended Data Fig. 5). Only the InsC–IAPP HIP (LQTLAL–NAARD) was verified as a fused peptide (Table 3 and Fig. 4a). We did not identify the HIP InsC–ChgA (LQTLAL–WSRMD)<sup>37</sup>, corresponding to the epitope of the potent diabetogenic BDC2.5 CD4<sup>+</sup> T cell, possibly due to the presence of this peptide at levels below the limit of detection for the mass spectrometry method used in this study. In brief, when following strict...
were calculated using an unpaired two-tailed Student's t-test. P values in Fig. 3a are from n = 10 mice. A mirror plot showing the match of the deamidated Ins1C:53–61E peptide identified in the islet MHC-II peptidome (top), along with the synthetic version (bottom). The dividing lines in the peptide sequence indicate sites of fragmentation. Each top tick denotes assigned y ions and the bottom ticks indicate b ions. Competitive binding assay showing the relative MHC-II-binding affinity of the native Ins1C:51–61 versus the deamidated Ins1C:51–61E and Ins1C:53–61E. Data (means ± s.e.m.) are from n = 4 independent experiments. *P = 0.0358, **P = 0.0062. P values for Ins1C:51–61E: *P = 0.0156; **P = 0.0004. The P values in f and g were calculated using an unpaired two-tailed Student’s t-test.

Table 3 | Putative MHC-II-bound HIP sequences obtained from the initial database search

| Source | Putative HIP sequence | Probable correct assignment | Match to synthetic standard |
|--------|-----------------------|----------------------------|-----------------------------|
| Islets + pLN | LQTLAL-NAARD | NA | Yes |
| Islets + pLN | PQVEQELGGSGPDLQ-APV | PQVEQELGGSGPDLQ(-H, O)LA | No |
| Islets | PQVEQELL-WSRMQDLAK | PQVEQELLGGSGPD(+Na)LQTLAL | No |
| pLN | PQVEQELGGGP-APVDN | PQVEQELGGSGPDLQ(-H, O) | No |
| pLN | PQVEQELL-GNP | PQVEQELGGGP | No |

One HIP sequence (top row) was verified by matching to a synthetic standard. For the four peptides in the bottom four rows, the spectra of the synthetic standards did not match the biological samples and a probable correct assignment is given. NA, not applicable.
criteria for the identification of fused peptides, only one major HIP was found in the MHC-II peptidome.

**Crinosomes but not DCGs contain free HIPS.** According to several of the proposed mechanisms for spliced peptide formation, crinosomes should be favorable for HIP formation. We reanalyzed published mass spectrometry data from crinosomes, DCGs and the secretome. The initial search identified 134 spectra for putative HIPs in crinosomes and 239 in the secretome, but none in DCGs (results not shown). We filtered the putative HIP assignments by either PEAKS PTM or SPIDER and then performed additional filtering by published criteria. These stringent criteria identified 20 putative HIPs from crinosomes and the secretome (Supplementary Table 5). The HIPs from the two sources were similar in composition, in terms of being either cis-fused InsC–InsC or trans-fused InsC–IAPP HIPs (Supplementary Table 5). Three HIP sequences common between the crinosomes and the secretome were verified by matching their spectra to those of the corresponding synthetic peptides (Fig. 4b–d). Multiple truncated forms of a HIP formed by InsC and IAPP (EVED–TPVRSGTNPQM), which was previously identified in a proteolytic digest of islets, were detected in the secretome (Fig. 4b and Supplementary Table 5). The other two HIPs (Fig. 4c,d) were not previously reported. None of the HIPs from crinosomes or the secretome were identified in the MHC-II peptidome (Table 3 and Supplementary Table 5). These observations indicate that fused peptides are formed in crinosomes (a site of proteolytic activity and high peptide concentrations) and were secreted with insulin.

**Discussion**

Here, we show that the initiation of diabetic autoimmunity depends on a limited number of CD4+ T cells that recognize peptides from proinsulin. The immunogenic peptides were not only found in islets, but also in the peripheral lymphoid tissues. In addition to the native peptides, our analysis identified an important PTM—a deamidation in peptides derived from Ins1C—which enhanced the MHC-II-binding affinity and the corresponding CD4+ T cell reactivity. We also documented an MHC-II-bound HIP in the islet and pLN peptidome and identified several others as free HIPs in the crinosomes, where the HIPs were formed.

Several features of InsB- and InsC-derived peptides contribute to their high immunogenicity. Both are from proinsulin—a major β cell protein. The peptides were displayed as sets varying in size, including long segments. Some of the long peptides harbored flanking residues that favored a strong interaction of the nine-amino-acid binding core with MHC-II, resulting in enhanced CD4+ T cell responses. The immunogenic peptides were also presented in peripheral lymphoid tissues, as a result of the exocytosis from crinosomes. We contend that a panoply of insulin peptides are generated in the β cell crinosomes and that they flow to islet and peripheral APCs where they are bound and presented. Such a pathway is supported by our findings and by previous studies examining a diabetogenic T cell specific for InsB:12–20 (the epitope examined here) (Fig. 4c,d). This mode of generation and presentation of peptides has two consequences. First, it generated unique T cell epitopes
that were not produced from the uptake and internal processing of insulin or proinsulin by the islet APCs (a process that involves editing of the internalized protein)\(^{17,25}\). Second, the peptides were exported to secondary lymphoid tissues where they contributed to the reactivity of T cells\(^1\). Peripheral presentation is important because the absence of peripheral lymph nodes markedly dampens the development of autoimmune diabetes\(^{24,25}\).

We note that the abundance of the MHC-II-bound peptides did not necessarily correlate with immunogenicity. The Ins2C peptide family had the highest abundance in the islet peptidome but generated few CD4\(^+\) T cell responses. Only a few Ins2C-derived peptides contained the immunogenic C-terminal segments, and none of these had deamidation that may enhance their MHC-II-binding affinity. Additionally, due to the expression of proinsulin-2 but not proinsulin-1 in the murine thymus, high selection pressure may be imposed on the Ins2C-reactive T cells. In summary, the Ins2C family provides an example of a highly displayed autoantigen in the target organ that is non-immunogenic.

We documented naturally deamidated Ins1C peptides bound to MHC-II. Deamidation improved binding to I-A\(^\beta\), although CD4\(^+\) T cells exclusively recognizing the deamidated epitope were not detected. Thus, such a PTM will be important in improving the interaction of the CD4\(^+\) T cells with the wild-type peptides, but not by giving rise to unique T cells that bypass conventional control mechanisms (an explanation frequently mentioned to explain the potential importance of peptides with PTMs). T cells have been reported to recognize deamidated sequences of peptides in T1D\(^{11,12}\). We identified deamidation in Ins1C:51–61 (LQTLALEV ARQ), but not in Ins2C:53–63 (LQTLALEVAQQ), despite similar abundance. This finding underscores the influence of the amino acids proximal to the glutamine in the deamidation process and suggests a non-enzymatic process\(^{11,25}\).

Although we identified the MHC-II-bound peptides involved at the initial stage of the autoimmune process in NOD mice, it is possible that we have missed peptides presented at a low abundance. Whether a very low amount of pMHCs in APCs can trigger CD4\(^+\) T cells is an issue to contend with. The finding that CD4\(^+\) T cells exclusively recognizing the deamidated epitope were not detected. Thus, such a PTM will be important in improving the interaction of the CD4\(^+\) T cells with the wild-type peptides, but not by giving rise to unique T cells that bypass conventional control mechanisms (an explanation frequently mentioned to explain the potential importance of peptides with PTMs). T cells have been reported to recognize deamidated sequences of peptides in T1D\(^{11,12}\). We identified deamidation in Ins1C:51–61 (LQTLALEV ARQ), but not in Ins2C:53–63 (LQTLALEVAQQ), despite similar abundance. This finding underscores the influence of the amino acids proximal to the glutamine in the deamidation process and suggests a non-enzymatic process\(^{11,25}\).

Two findings may have clinical relevance. First, examining the peptidome of the secondary lymphoid tissues draining a target organ may provide relevant information on the key autoantigens. This has clinical applicability if sampling the affected organs is not feasible, as it happens in pancreatic islets. Second, considering that I-A\(^\beta\), HLA-DQ2 and HLA-DQ8 are structural homologs\(^{2,3}\) and select similar peptides\(^6\), our analysis provides a foundation for assessing the T1D-relevant epitopes presented by them. We posit that the similar peptide families may be shared between the human and murine MHC-II peptidome. HLA-DQ8-restricted T cells recognize InsB9–23 (ref. \(^{20}\)) and the full-length C-peptide\(^6\) have been identified.

Finally, we confirmed the identity of an MHC-II-bound HIP and identified a few free fused peptides in crinosomes and in the secretome. Their identification in crinosomes—a structure rich in cathepsins—points to their formation via reverse proteolysis\(^{39}\). Our reported fused peptides were filtered in a very stringent manner to remove false positive assignments by eliminating all spectra that could be assigned to chemically or post-translationally modified forms.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-020-0623-7.

Received: 2 November 2019; Accepted: 3 February 2020; Published online: 9 March 2020

References
1. Hattori, M. et al. The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. Science 231, 733–735 (1986).
2. Todd, J. A., Bell, J. I. & McDevitt, H. O. HLA-DQ, gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. Nature 329, 599–604 (1987).
3. Acha-Orbea, H. & McDevitt, H. O. The first external domain of the nonobese diabetic mouse class II A-\(\beta\) chain is unique. Proc. Natl Acad. Sci. USA 84, 2435–2439 (1987).
4. Miyazaki, T. et al. Direct evidence for the contribution of the unique I-ANOD to the development of insulitis in non-obese diabetic mice. Nature 345, 722–724 (1990).
5. Corper, A. L. et al. A structural framework for deciphering the link between I-A\(^\beta\) and autoimmune diabetes. Science 288, 505–511 (2000).
6. Latек, R. R. et al. Structural basis of peptide binding and presentation by the type 1 diabetes-associated MHC class II molecule of NOD mice. Immunity 12, 699–710 (2000).
7. Lee, K. H., Wucherpfennig, K. W. & Wiley, D. C. Structure of a human insulin peptide–HLA-DQ8 complex and susceptibility to type 1 diabetes. Nature Immunol. 2, 501–507 (2001).
8. Suri, A., Walters, J. I., Gross, M. L. & Unanue, E. R. Natural peptides selected by diabetogenic DQ8 and murine I-A\(^\alpha\) molecules show common sequence specificity. J. Clin. Invest. 115, 2268–2276 (2005).
9. Lund, T. et al. Prevention of insulin-dependent diabetes mellitus in non-obese diabetic mice by transgenes encoding modified I-A\(^\beta\)-chain or normal I-E\(^\alpha\)-chain. Nature 345, 727–729 (1990).
10. Singer, S. M. et al. Prevention of diabetes in NOD mice by a mutated I-A\(^\beta\) transgene. Diabetes 47, 1570–1577 (1998).
11. Gioia, L. et al. Position J\(^{\beta}\) of I-A\(^\beta\) controls early anti-insulin responses in NOD mice, linking an MHC susceptibility allele to type 1 diabetes onset. Sci. Immunol. 4, eaaw6329 (2019).
12. Nakayama, M. et al. Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice. Nature 435, 220–223 (2005).
13. Daniel, D., Gill, R. G., Schloot, N. & Wegmann, D. Epitope specificity, cytokine production profile and diabetogenic activity of insulin-specific T cell clones isolated from NOD mice. J. Immunol. 155, 1056–1062 (1995).
14. Wegmann, D. R., Norbury- Glaser, M. & Daniel, D. Insulin-specific T cells are a predominant component of islet infiltrates in pre-diabetic NOD mice. Eur. J. Immunol. 24, 1853–1857 (1994).
15. Halbout, P., Briand, J.-P., Bécourt, C., Muller, S. & Boitard, C. T cell response to proinsulin I and II in the nonobese diabetic mouse. J. Immunol. 169, 2426–2443 (2002).
16. Levissetti, M. G., Lewis, D. M., Suri, A. & Unanue, E. R. Weak proinsulin peptide–major histocompatibility complexes are targeted in autoimmune diabetes in mice. Diabetes 57, 1852–1860 (2008).
17. Mohan, J. F. et al. Unique autoactive T cells recognize insulin peptides generated within the islets of Langherhans in autoimmune diabetes. Nat. Immunol. 11, 350–354 (2010).
18. Ziegler, A. G. et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. J. Am. Med. Assoc. 309, 2473–2479 (2013).
19. Rabon, J. A. B. et al. Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. Nat. Med. 22, 1482–1487 (2016).
20. Michels, A. W. et al. Islet-derived CD4 T cells targeting proinsulin in human autoimmune diabetes. Diabetes 66, 722–734 (2017).
21. Wan, X. et al. Pancreatic islets communicat e with lymphoid tissues via expression of insulin peptides. Nature 560, 107–111 (2018).
22. Unanue, E. R. Antigen presentation in the autoimmune diabetes of the NOD mouse. Annu. Rev. Immunol. 32, 579–608 (2014).
23. Unanue, E. R., Ferris, S. T. & Carrero, J. A. The role of islet antigen presenting cells and the presentation of insulin in the initiation of autoimmune diabetes in the NOD mouse. Immunol. Rev. 272, 183–201 (2016).
24. Andreatta, M., Alvarez, B. & Nielsen, M. G. Staff Cluster: unsupervised clustering and alignment of peptide sequences. Nucleic Acids Res. 45, W458–W463 (2017).
25. Mohan, J. F., Petzold, S. J. & Unanue, E. R. Register shifting of an insulin peptide–MHC complex allows diabetogenic T cells to escape thymic deletion. J. Exp. Med. 208, 2375–2383 (2011).

26. Wiles, T. A. et al. An insulin-IAPP hybrid peptide is an endogenous antigen for CD4 T cells in the non-obese diabetic mouse. J. Autoimmun. 78, 11–18 (2017).

27. Delong, T. et al. Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. Science 351, 711–714 (2016).

28. Gagnerault, M.-C., Luan, J. J., Lotton, C. & Lepault, F. Pancreatic lymph nodes are required for priming of β cell reactive T cells in NOD mice. J. Exp. Med. 196, 369–377 (2002).

29. Levisetti, M. G., Suri, A., Frederick, K. & Unanue, E. R. Absence of lymph nodes in NOD mice treated with lymphotoxin-β receptor immunoglobulin protects from diabetes. Diabetes 53, 3115–3119 (2004).

30. Faridi, P., Purcell, A. W. & Croft, N. P. In immunopeptidomics we need a sniper instead of a shotgun. Proteomics. 18, e1700464 (2018).

31. Babad, J., Geliebter, A. & DiLorenzo, T. P. T-cell autoantigens in the non-obese diabetic mouse model of autoimmune diabetes. Immunology 131, 459–465 (2010).

32. Mohan, J. F., Calderon, B., Anderson, M. S. & Unanue, E. R. Pathogenic CD4+ T cells recognizing an unstable peptide of insulin are directly recruited into islets bypassing local lymph nodes. J. Exp. Med. 210, 2403–2414 (2013).

33. Chang, K. Y. & Unanue, E. R. Prediction of HLA-DQ8 β cell peptidome using a computational program and its relationship to autoreactive T cells. Int. Immunol. 21, 705–713 (2009).

34. Aaron Wiles, T. et al. Identification of hybrid insulin peptides (HIPs) in mouse and human islets by mass spectrometry. J. Proteome Res. 18, 814–825 (2019).

35. Han, X., He, L., Xin, L., Shan, B. & Ma, B. PeaksPTM: mass spectrometry-based identification of peptides with unspecified modifications. J. Proteome Res. 10, 2930–2936 (2011).

36. Han, Y., Ma, B. & Zhang, K. SPIDER: software for protein identification from sequence tags with de novo sequencing error. J. Bioinform. Comput. Biol. 3, 697–716 (2005).

37. Berkers, C. R., de Jong, A., Ovaa, H. & Rodenko, B. Transpeptidation and reverse proteolysis and their consequences for immunity. Int. J. Biochem. Cell Biol. 41, 66–71 (2009).

38. Ebstein, F. et al. Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. Sci. Rep. 6, 24032 (2016).

39. Mishto, M. & Liepe, J. Post-translational peptide splicing and T cell responses. Trends Immunol. 38, 904–915 (2017).

40. Mohan, J. F. & Unanue, E. R. Unconventional recognition of peptides by T cells and the implications for autoimmunity. Nat. Rev. Immunol. 12, 721–728 (2012).

41. Van Lummel, M. et al. Posttranslational modification of HLA-DQ binding islet autoantigens in type 1 diabetes. Diabetes 63, 237–247 (2014).

42. Marre, M. L. et al. Modifying enzymes are elicited by ER stress, generating epitopes that are selectively recognized by CD4+ T cells in patients with type 1 diabetes. Diabetes 67, 1356–1368 (2018).

43. Wright, H. T. Nonenzymatic deamidation of asparaginyl and glutaminyl residues in proteins. Crit. Rev. Biochem. Mol. Biol. 26, 1–52 (1991).

44. Robinson, N. E. et al. Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. J. Pept. Res. 63, 426–436 (2004).

45. Vader, L. W. et al. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. J. Exp. Med. 195, 643–649 (2002).

46. Tisch, R. et al. Immune response to glutamic acid decarboxylase correlates with insulitis in non-obese diabetic mice. Science 366, 72–75 (1993).

47. Kelemen, K., Wegmann, D. R. & Hutton, J. C. T-cell epitope analysis on the autoimmune peptide–MHC complex allows diabetogenic T cells to escape thymic deletion. Proc. Natl Acad. Sci. USA 108, 1729–1734 (2001).

48. Tui, N. et al. Microbial antigen mimics activate diabetogenic CD8 T cells and the implications for autoimmunity. Trends Immunol. 1729–1734 (2011).

49. Hebbandi Nanjundappa, R. et al. A gut microbial mimic that hijacks diabetogenic autoreactivity to suppress colitis. Cell. 171, 655–667.e17 (2017).

50. So, M. et al. Proinsulin C-peptide is an autoantigen in people with type 1 diabetes. Proc. Natl Acad. Sci. USA 115, 10732–10737 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Articles Nature Immunology

Islet isolation. Mice were killed and the peritoneal cavity was opened to expose the common bile duct. The duct leading into the duodenum was clipped and, under a microscope, a syringe fitted with a 30-G needle was inserted into the duct close to the liver. Type XI collagenase (0.4 mg ml⁻¹; Sigma–Aldrich) in 5 ml Hank’s BSS (no Ca²⁺; Sigma–Aldrich) was injected to inflate the pancreas. The pancreas was removed and digested at 37°C for 15 min. After digestion, crude islets were washed and collected into a petri dish. Under the microscope, debris and undigested pancreas were removed and islets were handpicked.

Isolation of the MHCIi peptide bond. We followed our original procedure, with modifications. Cells isolated from islets, pLNs or spleens were suspended in lysis buffer (40 mM MEVA, 40 mM MEGA, 1 mM phenylmethylsulfonyl fluoride, 0.2 mM iodoacetamide, 20 µg ml⁻¹ leupeptin and Roche cOmplete Mini Protease cocktail in phosphate-buffered saline) and rocked for 1 h at 4°C. The cell lysate was spun in a centrifuge at 20,000 g for 30 min at 4°C. To eliminate non-specific binding of peptides, the supernatant was first incubated with polyclonal mouse immunoglobulin G (Bio X Cell; 1.5 mg antibody per sample) bound to Sepharose 4B at 4°C for 30 min (see Supplementary Table 1). A mixture of two G16H-II peptides was added and the whole material was then transferred to an Eppendorf tube containing phosphate-buffered saline washed with the anti-I-A⁺ antibody (G2A2.4.7; 1.5 mg per sample) and incubated at 4°C overnight. The I-A⁺-Sepharose was applied to a column and washed four times as follows: 10 ml 150 mM NaCl and 10 ml Tris (pH 7.4); 10 ml 400 mM NaCl and 20 mM Tris (pH 7.4); 10 ml 150 mM NaCl and 20 mM Tris (pH 7.4); and 10 ml 20 mM Tris (pH 8.0). Peptides were eluted with 10% acetic acid and dried with a SpeedVac. Eluted peptides were passed over a dialysis track with a 10-kDa cutoff to remove traces of remaining detergent and were cleaned using C18 Spin Columns from Thermo Fisher Scientific (Pierce).

Identification of effector T cells in islets and pLNs. Islets and pLNs were harvested from female NOD mice of various ages. Handpicked islets were dispersed using a non-enzymatic cell dispersion solution (Sigma–Aldrich) for single-cell suspensions. Total cells from the pLNs were isolated by digestion with Liberase (125 µg ml⁻¹; Roche) and DNAse (50 µg ml⁻¹; Roche). Cells from islet and pLN cells were mixed together and cultured in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal bovine serum at a concentration of 2 × 10⁶ cells per ml. The culture was supplemented with 20 U ml⁻¹ IL-2 and 1 µM of the peptide of interest (challenge). After incubation at 37°C for 7 d, total live cells were collected from the culture using Histopaque-1119 and then subject to a second round of amplification. Accordingly, live cells were added to a fresh culture containing 20 U ml⁻¹ IL-2 and 1 µM peptide and collected into a petri dish. Under the microscope, debris and undigested pLN was collected into a petri dish. The cell lysate was washed and digested at 37°C for 15 min. After digestion, crude islets were washed and collected into a petri dish. Under the microscope, debris and undigested pancreas were removed and islets were handpicked.

Competitive binding assay. We used a cell assay using the C3g7 cell line expressing a high amount of I-A⁺. Different amounts of a known peptide (GKVATTVHAGYK), well known to bind to I-A⁺, were used to compete against the binding to a HEL peptide (AMKRFIFGYNGRYSL) in the HEL peptide used, offered at a concentration of 200 pmol in 200 µl of 50% of its maximum binding, as estimated by the biological response against a known T cell (clone 10E11). The competitor and the HEL peptides were incubated with C3g7 at 37°C for 60 min. Cells were then spun down, washed and tested against 5 × 10⁶ 10E11 T cells for 24 h. The T cell responses were probed by the standard antigen presentation assay. Half-maximum inhibitory concentration values for competitor peptides were generated by fitting the data using GraphPad Prism software as a four-parameter inhibitor versus response curve.

Antigen presentation assay. The T cell responses to a given antigen were probed by IL-2 production measured by culturing purified T cells from the indicated time points (1, 2, 4, 6, 8 and 10 h), one aliquot of C3g7 was removed from the incubation, washed and cultured with the indicator T cells overnight.

Immunization and cloning selection. Eight-week-old male NOD mice were immunized with peptide or protein antigens (10 nmol) emulsified in Complete Freund’s Adjuvant (CFA; Difco) subcutaneously in the footpads of the hind legs. The mice were age and sex matched and were randomized and equally distributed into experimental groups. Power analysis was used to determine the sample size in the biological experiments. The sample size of the mass spectrometry experiments was determined in preliminary experiments in which doubling the sample size from 5 to 10 satisfied our confidence for the technical and biological variation.

Mass spectrometry data analysis. Data files were uploaded to PEAKS X (Bioinformatics Solutions) for processing, de novo sequencing and database searching. The sequences were searched against the UniProt Mouse database (downloaded 25 February 2019; 22,286 entries) or a custom, in-house database consisting of previously reported T1D autoantigens plus all of our identified I-A⁺ peptides. We searched with mass error tolerances of 10 ppm and 0.01 Da for parent and fragment, respectively, no enzyme specificity and oxidation (M), deamidation (NQ) and trioxidation (C) as variable modifications. FDR estimation was enabled. The results were filtered to remove any peptide with a score less than 2.0, and higher-energy collisional dissociation tandem mass spectrometry was triggered above a threshold of 2.0 × 10⁶, with quadrupole isolation (0.7 Da) at 30-K resolution and a collision energy of 30%. Dynamic exclusion was used (60 s), and monoisotopic precursor selection was on.
**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**
The I-A\(^d\) binding epitope prediction program will be available upon request. The mass spectrometry data are available via ProteomeXchange with identifiers PXD015408 (MHC-II peptidomes) and PXD015726 (reanalysis of crinosomes, DCGs and secretome).

**References**
51. Sofron, A., Ritz, D., Neri, D. & Fugmann, T. High-resolution analysis of the murine MHC class II immunopeptidome. *Eur. J. Immunol.* **46**, 319–328 (2016).

**Acknowledgements**
Members of the Unanue laboratory provided advice on many aspects of this project. We thank K. Frederick for maintaining the animal colony, P. Zakharov for assistance in prediction of the MHC-II epitopes and M. Gross for advice on mass spectrometry. This study is supported by grants from the National Institutes of Health (DK120340, DK058177 and AI114551), Juvenile Diabetes Research Foundation and Kilo Diabetes and Vascular Research Foundation.

**Author contributions**
E.R.U., C.F.L., X.W. and A.N.V. planned the experiments, interpreted the results and evaluated the data. X.W. and O.J.P. prepared the biological samples. A.N.V. isolated the MHC-II molecules. C.F.L. performed the mass spectrometry experiments and analyzed the data. A.N.V. and X.W. carried out the immunological experiments. A.V.C. contributed the germ-free mice used in the experiments and reviewed and examined the paper. X.W., C.F.L. and E.R.U. wrote the paper with input from all authors.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Extended data is available for this paper at https://doi.org/10.1038/s41590-020-0623-7.
Supplementary information is available for this paper at https://doi.org/10.1038/s41590-020-0623-7.
Correspondence and requests for materials should be addressed to C.F.L. or E.R.U.

**Peer review information** Ioana Visan was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Assessing the MHCII peptidomes of the pancreatic islets, pancreatic lymph nodes and spleens from NOD mice. a, Representative FACS plots showing the APC populations in the islets from 8-10-week old female NOD mice. The major APCs in the islet were the islet macrophages (CD45+CD11c+F4/80+) and dendritic cells (CD45+CD11c+F4/80−), along with a minor population of B cells (CD45+CD11c−F4/80−B220+). Data are representative of n = 4 independent experiment including n = 6 mice per experiment. b, Workflow for isolating the MHCII peptidome followed by mass spectrometry and immunological analyses. c, The lengths of all the peptides identified in the MHCII peptidomes from the indicated sites. Most of the peptides were 14-17 residues long, and a small number was longer than 25 residues. Data (mean) are from the mass spectrometry analysis of the MHCII peptidomes depicted in Supplementary Tables 1–3. d, Antigen presentation assay showing responses of an Ins1C-specific CD4+ T cell hybridoma to C3g7 APCs treated with or without chloroquine and pulsed with the full-length 29-residue Ins1C:33-61. The assay measures IL-2 production by the CD4+ T cell hybridoma during stimulation with the cognate antigen, assessed by the proliferation (3H incorporation) of the IL-2-depedent cell line CTLL-2 (see Methods). The Data (mean ± s.e.m.) are representative of n = 2 independent experiments with similar results.
Extended Data Fig. 2 | Evaluation of T cell reactivity and relative MHCII binding affinity in peptides selected from the MHCII peptidome. **a**, Pooled islet and pLN cells from 8-10-week old female NOD mice were challenged with each indicated peptide for two cycles, and ELISPOT assays were conducted to read IFN-γ production in the cells upon recalling with the same peptide (see Methods). The results show positive T cell responses to two immunogenic β-cell-derived peptides, Ins1C:33-61 and InsB:9-23, indicating the presence of the effector T cells to these peptides. No IFN-γ responses were observed to the HEL protein or the I-Ag7-binding HEL:11-25 peptide, demonstrating that the culture assay did not generate de novo T cells. Data (mean ± s.e.m.) are representative of n = 3 experiments using n = 5 mice per experiment. **b**, Different numbers of pre-activated HEL-reactive T cells were spiked into the normal two-cycle culture, followed by recall with the HEL peptide and ELISPOT assay. The data depict a ~1:1 recovery ratio of the spiked HEL-reactive T cells with the IL-2 spots. The assay detected as few as ~20 pre-activated HEL-reactive T cells, indicating a high sensitivity. Data (mean ± s.e.m.) are representative of n = 2 experiments using n = 8 mice per experiment. **c**, An example of the competitive binding assay and the presentation of the data. A standard APC line (C3g7) was cultured with 1µM HEL:11-25 peptide together with serial dilutions of a competitor peptide: the response of a specific T cell hybridoma to HEL:11-25 was probed by standard antigen presentation assay (see Methods). In every experiment, each competitor peptides were compared to a reference peptide, g7-MIME, with known strong binding to I-A^d^. The amounts of the peptide required to compete half-maximal T cell response to HEL:11-25 (IC50) was estimated and compared to the reference peptide. A higher amount indicated a weaker binding affinity. The table (right) depicts the IC50 calculated from several representative peptides. The results are presented as the percent of reference after normalization to the reference g7-MIME peptide. Data are representative of n = 4 independent experiments with similar results.
Extended Data Fig. 3 | CD4⁺ T cell recognition of the InsB:15-23 register is influenced by the nature of the flanking residues. a, Predicted I-A⁴⁺-binding registers included in the InsB:11-25 peptide. A preferred binding register is indicated by a log of odds (LOD) score (see Methods). b, Clone 58 also reacted with the InsB:12-26 peptide, an MHCII-bound sequence identified in the spleen peptidome. Mutation of the G23 into R23, an inhibitory residue, nullified the response, suggesting that the G23 was the P9 anchoring residue. Clone 58 is unreactive to InsB:12-20 or InsB:13-21. c, Comparison of T cell (clone 58) recognition between InsB:11-25 and peptides with varied residues flanking the InsB:15-23 register. The data showed the importance of the flanking residues in T cell recognition. The InsB:15-23 segment without any flanking residues did not induce any responses. Reducing, removing, or mutating the flanking residues at either the amino or carboxy end compromised the recognition to different extents. We note the role of having hydrophobic residues at the carboxy flank: changing the native FF residues into hydrophobic WW preserved the responses. Data (b,c) are representative of n=2 independent experiments with similar results.
Extended Data Fig. 4 | Analysis of the immunogenic segment in InsC peptides. a, ELISPOT assay showing IFN-γ and IL-2 production by CD4+ T cells from 8-week old male NOD mice immunized with two InsC peptides lacking the complete C-terminus upon recalling with indicated relevant peptides. Responses to either InsC:41-55 (left) or InsC:37-56 (right) were indistinguishable from background (no antigen); the positive control (ConA) generated strong responses. Data (mean ± s.e.m.) summarize results from n = 3 independent experiments from n = 6 mice. b, The mass spectrometry spectrum of the synthetic citrullinated InsC:53-61 peptide (TLALEVArQ; the lowercase r indicates the presence of citrulline). This spectrum was distinct from the deamidated InsC:53-61E peptide found in the islet MHCII peptidome, confirming that the biologically induced PTM was deamidation but not citrullination.
Extended Data Fig. 5 | An example of a false positive MHCII-bound HIP. Mass spectrum of a putative Ins1C-ChgA (PQVEQLEL-WSRMDQLAK) peptide in the islet MHCII peptidome (upper) that failed to match the synthetic standard peptide (middle). Further analysis suggested a probable correct match of the putative Ins1C-ChgA HIP to an Ins1C peptide fragment with sodium adduct (lower).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection
The T cell reactivity data were collected using the Cell counting software v 5.0 (ImmunoSpot). The flow cytometry data were collected using the FACS DIVA v 8.0.1 software (BD Biosciences). The mass spectrometry data were collected using XCalibur v 4.0 (ThermoFisher Scientific).

Data analysis
The biological data were analyzed using the Graphpad Prism 7.0. The flow cytometry data were analyzed using the Flowjo 10.0 software [Tree Star Software]. The mass spectrometry data were analyzed using PEAKS X [Bioinformatics Solutions]. The prediction of the I-Ag/ binding epitopes used a computational algorithm (TEA-DQ8) that was developed in the laboratory.

Policy information about availability of data

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The I-Ag/-binding epitope prediction program will be available upon request. The mass spectrometry data are available via ProteomeXchange with identifiers PXD015408 (MHCII peptides) and PXD015726 (reanalysis of crinosomes, DCGs and secretome).
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

For ELISPOT experiments, previous studies have shown that the mean numbers of the spots range from 20-50 with a standard deviation of ~10. To have a power of 95% and an alpha of 0.05, a minimal of 3 biological replicates are required. We involved 3-7 biological replicates in the studies. The sample size of mass spectrometry experiments was determined in preliminary experiments in which different sample sizes were tested to reach a satisfactory level of peptide recovery and detection.

Data exclusions

We did not exclude any data in our studies.

Replication

For biological experiments, the results were pooled from 2-5 independent experiments containing 1-4 biological replicates per experiment. Each biological replicates used materials from 2-5 mice. For mass spectrometry experiments that required materials from 4-5 mice per run, we summarized the results from 4 independent runs. For mass spectrometry experiments that required a large quantity of materials, we collected biological samples from 200 plus mice and the results were from one independent run with quality controls. All the replications in the biological experiments were consistent and showed the same trend of differences between different conditions; the variations among different experiments are noted by each data point in the figures. The mass spectrometry analysis of the spleen MHCII peptidome showed similar beta-cell-derived peptide families among four independent runs and the results of the individual runs are provided in Supplementary Table 3.

Randomization

Mice were age and gender matched; they were randomized and distributed equally into experimental groups.

Blinding

The investigators were not blinded to group allocation during data collection or analysis. All the measurements were generated by computational programs and algorithms without involving subjective scoring. Therefore, blinding is not necessary.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑️  | Antibodies            |
| ☑️  | Eukaryotic cell lines |
| ☑️  | Palaeontology         |
| ☑️  | Animals and other organisms |
| ☑️  | Human research participants |
| ☑️  | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☑️  | ChIP-seq              |
| ☑️  | Flow cytometry        |
| ☑️  | MRI-based neuroimaging |

Antibodies

The anti-mouse I-Ag7 antibody (clone AG2.42.7) was generated and produced in BioXcell. The polyclonal mouse IgG (cat#: B0093; lot: 63841601) was purchased from BioLegend. The following fluorescently conjugated anti-mouse antibodies were purchased from BioLegend: PerCP/CD5.5-anti-B220 (clone: RA3-6B2; cat#: 103836; lot: B2531647; dilution: 1:200); PE/Cy7-anti-CD11c (clone: N418; cat#: 117318; lot: B283053; dilution: 1:200); Brilliant Violett-anti-CD45 (clone: 30-F11; cat#: 103138; lot: B296483; dilution: 1:200); APC-anti-F4/80 (clone: BM8; cat#: 122313; lot: B781628; dilution: 1:200); and PE-anti-Thy1.2 (clone: 30-H12; cat#: 140309; lot: B148622; dilution: 1:400).

Validation

The reactivity of the anti-I-Ag7 antibody was confirmed by flow cytometry experiments showing that 95-100% of the C3g7 APCs were stained positive by the Pacific Blue conjugated AG2.42.7 antibody; whereas less than 1% cells of a I-Ag7-negative control cell line were stained positive. Other commercial antibodies were validated by the vendor. Detailed information can be found at: BioXcell; https://bxcell.com/invivo-mab-vs-invivo-plus/ Biolegend;
Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)  The CD4+ T cell hybridomas and the C3g7 B cell lymphoma line (used as APCs) were generated in the laboratory.

Authentication  The reactivity of each CD4+ T cell hybridoma to their specific antigens were tested and confirmed by their abilities to produce interleukin-2 upon stimulation with antigen presentation cells pulsed with the cognate antigens. Limited dilution followed by sub-cloning for 2-4 rounds was used for each CD4+ T cell hybridoma to generate monoclonal T cell clones.

Mycoplasma contamination  All the cell lines were confirmed as mycoplasma free.

Commonly misidentified lines (See ICLAC register)  Not commonly misidentified lines were used in the study.

Animals and other organisms

Policy information about studies involving animals: ARRIVE guidelines recommended for reporting animal research

Laboratory animals  NOD/ShiLtJ (NOD), NOD.129S5(B6)-Rag1tm1Mom/J (NOD.Rag1−/−), NOD.129S2(B6)-Ighmtrm1Cgn/D0j (μMT), and NOD.B10Sn-H2b/J [NOD.H2b] mice were originally obtained from the Jackson Laboratory. NOD.1DE11 TCR transgenic mice to HEL were generated in the laboratory. Germ-free NOD mice were provided by A. Chervonsky. For most of the experiments, including mass spectrometry, in vivo ELISPOT assays, and adoptive transfer, 8-10-week old female mice were used unless otherwise noted. For immunization experiments, 8-week old male mice were used.

Wild animals  No wild animals were used in the study.

Field-collected samples  No field-collected samples were used in the study.

Ethics oversight  All experiments were approved by the Division of Comparative Medicine of Washington University School of Medicine in St. Louis (Accreditation number A3381-01).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation  Single cell suspensions were prepared. Cells were then incubated with FcR blocking media containing 2.4G2 antibody for 30 min on ice. The cells were then stained for CD4, CD8, B220, CD11c, and CD44 with corresponding fluorescent antibodies on ice for 25 min.

Instrument  FACSCanto II (BD Biosciences)

Software  Data were collected using FACS DIVA v 8.0.1 (BD Biosciences) and were analyzed using FlowJo v.10.00 software (Tree Star Software).

Cell population abundance  Cell sorting was not used in the study.

Gating strategy  The islet macrophages were gated as singlets, live, CD45+, F4/80+, CD11c−; the islet-infiltrating dendritic cells were gated as singlets, live, CD45+, F4/80−, CD11c+; the islet-infiltrating T cells were gated as singlets, live, CD45+, F4/80+, CD11c−-B220−-thy1.2−; the islet-infiltrating B cells were gated as singlets, live, CD45+, F4/80+, CD11c+, B220+, thy1.2−.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary information.