

**Supplemental Figure 1. HIP-Database Generation.** HIP sequences in our database were generated through the linear combination of 86 proinsulin sequences on the N-terminal (left) side and 89 natural peptide sequences on the C-terminal (right) side. In the resulting HIPs, the C-terminal carboxylic acid residues (COOH) of the left peptides were linked to the N-terminal amino residues (H2N) of the right peptides. Each of the 86 insulin peptides (left peptides) contain 8 residues, with the exception of 7 peptides describing the N-terminal region of proinsulin. These peptides contain between 1-7 residues (F, FV, FVN, FVNQ, … , FVNQHLC). Each of the 89 natural peptides (right peptides) were derived from various proteins (Proinsulin, Secretogranin-1, Secretogranin-2, Secretogranin-3, Secretogranin-5, Chromogranin A, proSAAS, Neuroendocrine Convertase 2, Neuropeptide Y, Islet Amyloid Polypeptide, Insulin Like Growth Factor II, GRP78) that are present in islets and/or beta cell granule extracts. Each of these right-side peptides also contained 8 amino acid residues. The N-termini of these peptides were naturally preceded either by the proteins’ N-termini, or by dibasic amino acid residues (KK, KR, RK, RR), which are predicted cleavage sites of granular prohormone convertases. The resulting total number of HIP sequences in our database is 7654 (86 × 89).

****

**Supplemental Figure 2. Assessing the *in vitro* Immunogenicity of candidate HIP peptides.** To assess the immunogenicity of the candidate HIP sequences that bound to DR0401, PBMC from 8 unique subjects with type 1 diabetes (T1D) - all with DR0401 haplotypes - were expanded by stimulating for two weeks with groups of HIP peptides and then stained with corresponding individual HIP peptide loaded tetramers. Each staining panel above shows representative positive (HIP-4.1, HIP-4.2, HIP-4.4, HIP-4.18, HIP-4.39, and HIP-4.40) or negative (HIP-4.7 and HIP-4.47) tetramer staining results, displayed as CD4 versus tetramer after gating for CD3+ lymphocytes.



**Supplemental Figure 3. Correlations between HIP reactive T cell frequency and clinical characteristics.** To draw inferences about the potential sources of variation in the frequencies of HIP reactive T cells, we performed regression and between group comparisons. (A) There was a significant negative correlation (p=0.0048, simple linear regression) between age and the combined frequency of HIP-reactive T cells. (B) The two subjects who were not IAA positive had significantly lower frequencies of HIP-reactive T cells than the remaining subjects (p=0.0173, Mann Whitney test), who were all IAA-positive.



**Supplemental Figure 4. Cytokine profiling of HIP reactive T cell clones.** To characterize their cytokine profiles, tetramer positive T cell clones representing each HIP specificity were activated using PMA/ionomycin and accumulation of IL-4, TNF-α, IFN-γ and IL-17 was determined by intracellular cytokine staining. (A) Staining result for a representative HIP-4.40 reactive T cell clone. Each panel shows tetramer versus cytokine staining (pre-gated for CD4+ lymphocytes). (B) Clones that recognized different HIP peptides exhibited some degree of variation in their cytokine profiles, but in general HIP-reactive clones had comparatively higher levels of IFN-ɣ than TNF-α or IL4 and produced little to no IL-17.

****

**Supplemental Figure 5. Phenotype of CD4+ T cells specific for novel HIPs in healthy controls.** (**A**) In contrast to subjects with T1D, across all specificities, the phenotype of HIP-specific CD4+ T cells in controls was evenly distributed between memory and naïve. (**B**) Among CD45RA- T cells, HIP-specific CD4+ T cells in controls were more central memory-like (CCR7+) as opposed to being effector-like (CCR7-).

**Supplemental Table 1. DRB1\*4:01+ subjects with diabetes**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Subject ID** | **Agea** | **Sex** | **HLA-DR Typeb** | **Time Since Diagnosisa** | **Auto-antibody statusc** |
| T1D #1 | 36 | Female | \*04:01/\*04:01 | 31.5 years | IAA, IA2 |
| T1D #2 | 34 | Female | \*04:01/\*03:01 | 31.6 years | IAA |
| T1D #3 | 39 | Female | \*04:01/\*01:01 | 27.9 years | GAD, IA2 |
| T1D #4 | 34 | Male | \*04:01/\*10:01 | 7.0 years | GAD, IAA |
| T1D #5 | 32 | Male | \*04:01/\*13- | 6.2 years | GAD, IAA, IA2, ZnT8 |
| T1D #6 | 37 | Male | \*04:01/\*03- | 1.8 years | IAA, ZnT8 |
| T1D #7 | 33 | Male | \*04:01/\*03- | 2.6 years | GAD, IAA, ZnT8 |
| T1D #8 | 32 | Female | \*04:01/\*01:01 | 20.8 years | GAD, IAA, IA2 |
| T1D #9 | 25 | Male | \*04:01/\*03:01 | 4.9 years | GAD, IAA, IA2, ZnT8 |
| T1D #10 | 32 | Female | \*04:01/\*03:01 | 15.6 years | GAD, IAA |
| T1D #11 | 31 | Male | \*04:01/\*03- | 3.9 years | GAD, IAA |
| T1D #12 | 17 | Male | \*04:01/ND | 9.4 years | IAA |
| T1D #13 | 27 | Male | \*04:01/\*04:04 | 1.4 years | GAD, IAA |
| T1D #14 | 30 | Female | \*04:01/\*15:02 | 4.0 years | GAD, IA2, ZnT8 |
| T1D #15 | 28 | Male | \*04:01/\*13- | 3.8 years | GAD, IAA, IA2, ZnT8 |
| T1D #16 | 31 | Female | \*04:01/\*03:01 | 3.8 years | GAD, IAA |
| T1D #17 | 21 | Female | \*04:01/\*03:01 | 3.2 years | GAD, IAA, IA2, ZnT8 |
| T1D #18 | 27 | Female | \*04:01/\*04:01 | 2.4 years | GAD, IAA, IA2 |
| T1D #19 | 17 | Female | \*04:01/\*03- | 4.6 years | GAD, IAA, IA2 |
| T1D #20 | 34 | Male | \*04:01/\*ND | 4.3 years | GAD, IAA, IA2, ZnT8 |
| T1D #21 | 14 | Male | \*04:01/\*ND | 1.0 years | GAD, IAA, IA2, ZnT8 |
| T1D #22 | 16 | Male | \*04:01/\*01:01 | 4.5 years | IAA, IA2 |

a The mean age of patients was 28.5 years. The average time since diagnosis was 11.2 years

b ND indicates that the DR type could not be determined by the method used, dash mark (-) indicates that the high resolution DR type was not conclusive

c GAD denotes glutamic acid decarboxylase 65, IAA denotes (micro) insulin, IA2 denotes tyrosine phosphatase-related islet antigen 2, ZnT8 denotes zinc transporter 8

**Supplemental Table 2. DRB1\*04:01+ healthy subjects**

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject ID** | **Agea** | **Sex** | **HLA-DR Typeb** |
| Control #1 | 21 | Female | \*04:01/\*09:01 |
| Control #2 | 23 | Female | \*04:01/\*13:02 |
| Control #3 | 26 | Male | \*04:01/\*01:01 |
| Control #4 | 22 | Male | \*04:01/\*15:01 |
| Control #5 | 27 | Male | \*04:01/\*01:01 |
| Control #6 | 25 | Female | \*04:01/\*15:01 |
| Control #7 | 24 | Male | \*04:01/\*12:01 |
| Control #8 | 23 | Female | \*04:01/\*04:01 |
| Control #9 | 46 | Male | \*04:01/\*03- |
| Control #10 | 57 | Female | \*04:01/\*04:01 |
| Control #11 | 36 | Male | \*04:01/\*01:01 |
| Control #12 | 36 | Female | \*04:01/\*12:01 |
| Control #13 | 64 | Female | \*04:01/\*15:01 |
| Control #14 | 32 | Female | \*04:01/\*03:01 |
| Control #15 | 25 | Female | \*04:01/\*15:01 |
| Control #16 | 28 | Female | \*04:01/\*04:03 |
| Control #17 | 27 | Female | \*04:01/\*13:01 |
| Control #18 | 24 | Male | \*04:01/\*13- |
| Control #19 | 46 | Male | \*04:01/\*07:01 |
| Control #20 | 55 | Male | \*04:01/\*13:02 |
| Control #21 | 28 | Male | \*04:01/\*13- |

a The mean age of controls was 33.1 years. Healthy subjects under the age of 18 could not be recruited

b Dash mark (-) indicates that the high resolution DR type was not conclusive

**Supplemental Table 3. Prediction Matrix for binding to DRB1\*04:01**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **P1**a  **residue** | **Cp1**b | **P4**a  **residue** | **Cp4**b | **P6**a  **residue** | **Cp6**b | **P7**a  **residue** | **Cp7**b | **P9**a  **residue** | **Cp9**b |
| G | 0.00 | G | 0.31 | G | 0.43 | G | 1.80 | G | 1.00 |
| A | 0.08 | A | 0.86 | A | 0.93 | A | 1.82 | A | 1.00 |
| V | 0.24 | V | 0.67 | V | 1.56 | V | 0.60 | V | 0.50 |
| L | 0.24 | L | 1.32 | L | 0.65 | L | 0.42 | L | 0.40 |
| I | 0.30 | I | 1.16 | I | 0.91 | I | 0.42 | I | 0.35 |
| M | 0.18 | M | 2.08 | M | 0.54 | M | 1.08 | M | 1.12 |
| P | 0.00 | P | 0.07 | P | 0.90 | P | 0.10 | P | 0.17 |
| F | 1.50 | F | 1.97 | F | 0.30 | F | 0.51 | F | 0.10 |
| W | 0.93 | W | 0.47 | W | 0.39 | W | 0.60 | W | 0.10 |
| S | 0.00 | S | 1.14 | S | 1.36 | S | 0.57 | S | 3.48 |
| T | 0.00 | T | 0.91 | T | 1.83 | T | 0.57 | T | 0.85 |
| N | 0.00 | N | 3.65 | N | 1.21 | N | 1.50 | N | 0.72 |
| Q | 0.00 | Q | 1.04 | Q | 0.14 | Q | 0.90 | Q | 0.60 |
| Y | 0.97 | Y | 0.89 | Y | 0.27 | Y | 0.56 | Y | 0.18 |
| C | 0.00 | C | 1.36 | C | 0.72 | C | 0.12 | C | 0.40 |
| K | 0.00 | K | 0.06 | K | 0.02 | K | 0.03 | K | 0.05 |
| R | 0.00 | R | 0.03 | R | 0.02 | R | 0.09 | R | 0.07 |
| H | 0.00 | H | 1.27 | H | 0.30 | H | 0.39 | H | 2.64 |
| D | 0.00 | D | 1.17 | D | 0.84 | D | 0.12 | D | 0.70 |
| E | 0.00 | E | 1.46 | E | 1.17 | E | 0.12 | E | 0.40 |

a P1, P4, P6, P7, and P9 denote binding pocket 1, binding pocket 4, binding pocket 6, binding pocket 7, and binding pocket 9 of HLA-DR0401

b Cp1,  Cp4, etc. denotes the predicted influence of each residue on binding to the corresponding pocket. The relative binding affinity of any peptide is predicted as Cp1×Cp4×Cp6×Cp7×Cp9

**Supplemental Table 4. Top 50 Non-Redundant Theoretical DR4 HIPs**

|  |  |  |  |
| --- | --- | --- | --- |
| Peptide | Sequence | Molcular Weight (g/mol) | IC50  (µM)a,b |
| **HIP-4.1** | **VCGERGFFEELVARSE** | **1828** | **2.8** |
| **HIP-4.2** | **HLVEALYLEELVARSE** | **1871** | **0.44** |
| **HIP-4.3** | **TSICSLYQEELVARSE** | **1828** | **1.3** |
| **HIP-4.4** | **ICSLYQLEFVNQHLCG** | **1867** | **10.3** |
| **HIP-4.5** | **FVNQFVNQHLCG** | **1405** | **33.9** |
| **HIP-4.6** | **GERGFFYTFVNQHLCG** | **1875** | **4.5** |
| **HIP-4.7** | **CSLYQLENAADHDVGS** | **1722** | **0.41** |
| HIP-4.8 | VCGERGFFFLGEGHHR | 1848 | >50 |
| HIP-4.9 | CGERGFFYFLGEGHHR | 1912 | >50 |
| HIP-4.10 | VEALYLVCFVNQHLCG | 1808 | >50 |
| **HIP-4.11** | **ALYLVCGEALSSQHQA** | **1690** | **24.2** |
| HIP-4.12 | FVNQHLALSSQHQA | 1580 | >50 |
| HIP-4.13 | CGERGFFYNGVGLEFN | 1809 | >50 |
| **HIP-4.14** | **GERGFFYTNGVGLEFN** | **1807** | **20.9** |
| HIP-4.15 | ERGFFYTPAADHDVGS | 1769 | >50 |
| **HIP-4.16** | **CGERGFFYLVNAAGSG** | **1648** | **4.6** |
| HIP-4.17 | EALYLVCGAADHDVGS | 1620 | >50 |
| **HIP-4.18** | **SLQKRGIVEELVARSE** | **1814** | **1.8** |
| **HIP-4.19** | **EQCCTSICEELVARSE** | **1800** | **0.41** |
| **HIP-4.20** | **CSLYQLENNQLHDEVH** | **1942** | **32.9** |
| **HIP-4.21** | **HLVEALYLFLGEGHHR** | **1891** | **1.0** |
| HIP-4.22 | TSICSLYQFLGEGHHR | 1848 | >50 |
| HIP-4.23 | FVNEELVARSE | 1292 | >50 |
| **HIP-4.24** | **HLCGSHLVEELVARSE** | **1779** | **1.5** |
| **HIP-4.25** | **SHLVEALYEELVARSE** | **1845** | **3.0** |
| HIP-4.26 | RREAEDLQEELVARSE | 1930 | >50 |
| HIP-4.27 | DLQVGQVEEELVARSE | 1801 | >50 |
| HIP-4.28 | VCGERGFFESKDQLSD | 1817 | >50 |
| HIP-4.29 | CGERGFFYEEEGSAN | 1695 | >50 |
| HIP-4.30 | FVNQHNQLHDEVH | 1616 | >50 |
| **HIP-4.31** | **ICSLYQLELVNAAGSG** | **1638** | **0.06** |
| **HIP-4.32** | **GERGFFYTLVNAAGSG** | **1646** | **0.09** |
| **HIP-4.33** | **CSLYQLENNYPSLELD** | **1901** | **0.10** |
| HIP-4.34 | CGERGFFYNPEAGVAT | 1718 | >50 |
| HIP-4.35 | ALYLVCGEMRSAAVLA | 1667 | >50 |
| **HIP-4.36** | **ICSLYQLEKTGKDV** | **1597** | **11.4** |
| **HIP-4.37** | **CGERGFFYEDVGTVVG** | **1735** | **28.2** |
| **HIP-4.38** | **CSLYQLENNGVGLEFN** | **1800** | **0.10** |
| **HIP-4.39** | **CSLYQLENSVPHFSDE** | **1868** | **0.24** |
| **HIP-4.40** | **QPLALEGSALSSQHQA** | **1637** | **2.4** |
| HIP-4.41 | LVEALYLVNGVGLEFN | 1750 | >50 |
| **HIP-4.42** | **ERGFFYTPNGVGLEFN** | **1847** | **0.23** |
| HIP-4.43 | QPLALEGSGESRSEAL | 1644 | >50 |
| **HIP-4.44** | **ERGFFYTPNQLHDEVH** | **1989** | **2.4** |
| **HIP-4.45** | **CGERGFFYMRSAAVLA** | **1778** | **0.49** |
| **HIP-4.46** | **HLVEALYLLVNAAGSG** | **1627** | **23.6** |
| **HIP-4.47** | **GERGFFYTELENLAAM** | **1848** | **0.06** |
| **HIP-4.48** | **EALYLVCGNQLHDEVH** | **1840** | **1.1** |
| HIP-4.49 | GIVEQCCTGESRSEAL | 1682 | >50 |
| HIP-4.50 | ALYLVCGEQASAI | 1337 | >50 |

a IC50 represents the peptide concentration that displaces half of the reference peptide

b Peptides selected for in vitro studies (based on IC50 values) are shown in boldface

**Supplemental Table 5. Lack of binding for Half-HIP sequences**

|  |  |  |
| --- | --- | --- |
| **Peptide** | **Amino acid sequencea,b** | **IC50 (µM)c,d** |
| HIP-4.1  HIP-4.1L  HIP-4.1R | **VCGERGFF**EELVARSE  **VCGERGFF**YTPKTRRE  KQASAIKKEELVARSE | 2.0  >50  >50 |
| HIP-4.2  HIP-4.2L  HIP-4.2R | **HLVEALYL**EELVARSE  **HLVEALYL**VCGERGFF  KQASAIKKEELVARSE | 0.72  >50  >50 |
| HIP-4.4  HIP-4.4L  HIP-4.4R | **ICSLYQLE**FVNQHLCG  **ICSLYQLE**NYCN .  WGPDPAAAFVNQHLCG | 9.2  >50  >50 |
| HIP-4.18  HIP-4.18L  HIP-4.18R | **SLQKRGIVE**ELVARSE  **SLQKRGIVE**QCCTSIC  KQASAIKKEELVARSE | 3.8  >50  >50 |
| HIP-4.39  HIP-4.39L  HIP-4.39R | **CSLYQLEN**SVPHFSDE  **CSLYQLEN**YCN .  LDNVVAKKSVPHFSDE | 0.53  >50  >50 |
| HIP-4.40  HIP-4.40L  HIP-4.40R | **QPLALEGS**ALSSQHQA  **QPLALEGS**LQKRGIVE  REVEKAKRALSSQHQA | 2.9  >50  >50 |

aThe left peptide is bolded and the right is underlined in each sequence

bIdentical Half-HIP sequences were synthesized once but appear in multiple groups for the sake of clarity - dots indicate the C-terminus of the peptide

cIC50 represents the peptide concentration that displaces half of the reference peptide - representative values from two replicate experiments are shown

dPeptides with no detectable binding were assigned an IC50 >50 μM (limit of detection)

**Supplemental Table 6. Binding Arginine Substituted HIP sequences**

|  |  |  |
| --- | --- | --- |
| **Peptidea** | **Amino acid sequenceb,c** | **IC50 (µM)d,e** |
| HIP-4.1  HIP-4.1 R1  HIP-4.1 R7**a**  HIP-4.1 R8 | VCGERGFFEELVARSE  **R**CGERGFFEELVARSE  VCGERG**R**FEELVARSE VCGERGF**R**EELVARSE | 2.0  1.2  24  1.4 |
| HIP-4.2  HIP-4.2 R2  HIP-4.2 R3  HIP-4.2 R6  HIP-4.2 R7**a**  HIP-4.2 R8 | HLVEALYLEELVARSE  H**R**VEALYLEELVARSE  HL**R**EALYLEELVARSE HLVEA**R**YLEELVARSE HLVEAL**R**LEELVARSE HLVEALY**R**EELVARSE | 0.72  0.23  0.22  0.25  35  0.19 |
| HIP-4.4  HIP-4.4 R1  HIP-4.4 R4  HIP-4.4 R5**a**  HIP-4.4 R7 | ICSLYQLEFVNQHLCG  **R**CSLYQLEFVNQHLCG  ICS**R**YQLEFVNQHLCG  ICSL**R**QLEFVNQHLCG  ICSLYQ**R**EFVNQHLCG | 9.2  11  9.7  >50  16 |
| HIP-4.18  HIP-4.18 R2  HIP-4.18 R7**a**  HIP-4.18 R8 | SLQKRGIVEELVARSE  S**R**QKRGIVEELVARSE  SLQKRG**R**VEELVARSE  SLQKRGI**R**EELVARSE | 3.8  3.2  >50  1.4 |
| HIP-4.39  HIP-4.39 R3  HIP-4.39 R4**a**  HIP-4.39 R6 | CSLYQLENSVPHFSDE  CS**R**YQLENSVPHFSDE  CSL**R**QLENSVPHFSDE  CSLYQ**R**ENSVPHFSDE | 0.53  0.31  >50  0.59 |
| HIP-4.40  HIP-4.40 R3**a**  HIP-4.40 R5 | QPLALEGSALSSQHQA  QP**R**ALEGSALSSQHQA  QPLA**R**EGSALSSQHQA | 2.9  9.13  2.4 |

aIndicates substitutions that disrupt binding

bThe substituted arginine residue is bolded in each sequence

cThe minimal motif that is consistent with the effect of arginine substitution is underlined within the non-substituted peptide

dIC50 represents the peptide concentration that displaces half of the reference peptide. Representative values from two replicate experiments are shown

ePeptides with no detectable binding were assigned an IC50 >50 μM (limit of detection)