

A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides^{*[5]}

Received for publication, August 3, 2011, and in revised form, November 14, 2011. Published, JBC Papers in Press, November 18, 2011, DOI 10.1074/jbc.M111.289488

Linda Wooldridge^{†1,2}, Julia Ekeruche-Makinde^{†1}, Hugo A. van den Berg^{§1}, Anna Skowera^{¶1,3}, John J. Miles^{‡4}, Mai Ping Tan[‡], Garry Dolton^{‡,3}, Mathew Clement[‡], Sian Llewellyn-Lacey[‡], David A. Price^{‡,3}, Mark Peakman^{¶1,3}, and Andrew K. Sewell^{‡,3,5}

From the [†]Institute of Infection and Immunity, Cardiff University School of Medicine, Henry Wellcome Building, Heath Park, Cardiff CF14 4XN, United Kingdom, the [§]Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom, the [¶]Department of Immunobiology, King's College London, London SE1 9RT, United Kingdom, and the ^{||}National Institute for Health Research Biomedical Research Centre at Guy's and St. Thomas' National Health Service Foundation Trust and King's College London, London SE1 9RT, United Kingdom

Background: How does a limited pool of $<10^8$ T cell receptors (TCRs) provide immunity to $>10^{15}$ antigens?

Results: A single TCR can respond to $>$ one million different decamer peptides.

Conclusion: This unprecedented level of receptor promiscuity explains how the naïve TCR repertoire achieves effective immunity.

Significance: TCR degeneracy has enormous potential to be the root cause of autoimmune disease.

The T cell receptor (TCR) orchestrates immune responses by binding to foreign peptides presented at the cell surface in the context of major histocompatibility complex (MHC) molecules. Effective immunity requires that all possible foreign peptide-MHC molecules are recognized or risks leaving holes in immune coverage that pathogens could quickly evolve to exploit. It is unclear how a limited pool of $<10^8$ human TCRs can successfully provide immunity to the vast array of possible different peptides that could be produced from 20 proteogenic amino acids and presented by self-MHC molecules ($>10^{15}$ distinct peptide-MHCs). One possibility is that T cell immunity incorporates an extremely high level of receptor degeneracy, enabling each TCR to recognize multiple peptides. However, the extent of such TCR degeneracy has never been fully quantified. Here, we perform a comprehensive experimental and mathematical analysis to reveal that a single patient-derived autoimmune CD8⁺ T cell clone of pathogenic relevance in human type I diabetes recognizes $>$ one million distinct decamer peptides in the context of a single MHC class I molecule. A large number of peptides that acted as substantially better agonists than the wild-type “index” preproinsulin-derived peptide (ALWGPDPAAA) were

identified. The RQFGPDFPTI peptide (sampled from $>10^8$ peptides) was >100 -fold more potent than the index peptide despite differing from this sequence at 7 of 10 positions. Quantification of this previously unappreciated high level of CD8⁺ T cell cross-reactivity represents an important step toward understanding the system requirements for adaptive immunity and highlights the enormous potential of TCR degeneracy to be the causative factor in autoimmune disease.

The mammalian T cell receptor (TCR)⁶ orchestrates immune responses by binding to foreign peptides presented at the cell surface in the context of major histocompatibility complex (MHC) molecules. Recognition is mediated by the highly variable complementarity-determining regions of the $\alpha\beta$ TCR (1, 2). *A priori*, the TCR repertoire must be broad enough to respond to all foreign peptides that can bind to self-MHC molecules (3). If this were not the case, then pathogens could rapidly evolve to exploit such deficiencies in immune coverage. Current estimates of human $\alpha\beta$ TCR diversity suggest that there are $<10^8$ different antigen receptors in the naïve T cell pool (4), a number that is dwarfed by the potential number of antigenic peptide-MHC molecules that could be encountered. Although next generation sequencing technologies may lead to an increased estimate of TCR diversity, such future revisions are unlikely to alter the fact that a relatively small number of TCRs must, and do, provide effective immune recognition of all peptides that can be generated from 20 proteogenic amino acids and that also bind self-MHC molecules ($>10^{15}$ distinct peptide-MHCs). This represents a particular biochemical challenge to the immune system because the TCR, unlike the B cell receptor, cannot undergo affinity maturation in the form of somatic hypermutation.

* This work was supported by the Biotechnology and Biological Sciences Research Council Grant BB/H001085/1 and the United Kingdom Department of Health via the National Institute for Health Research comprehensive Biomedical Research Centre award to Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London (to A. S. and M. P.).

Author's Choice—Final version full access.

[5] This article contains supplemental Figs. S1–S6, Tables S1–S3, Equations S1 and S2, Methods, and additional references.

¹ These authors contributed equally to this work.

² Supported by Wellcome Trust Clinical Intermediate Fellow Grant WT079848MA.

³ Supported by Juvenile Diabetes Research Foundation Grants 7-2005-877, 1-2007-1803, or 17-2009-806.

⁴ Supported by a Welsh Office of Research and Development translational fellowship.

⁵ To whom correspondence should be addressed. Tel.: 44-29-206-87055; Fax: 44-29-206-87007; E-mail: sewellak@cardiff.ac.uk.

⁶ The abbreviations used are: TCR, T cell receptor; CPL, combinatorial peptide library; HLA, human leukocyte antigen; MIP, macrophage inflammatory protein.

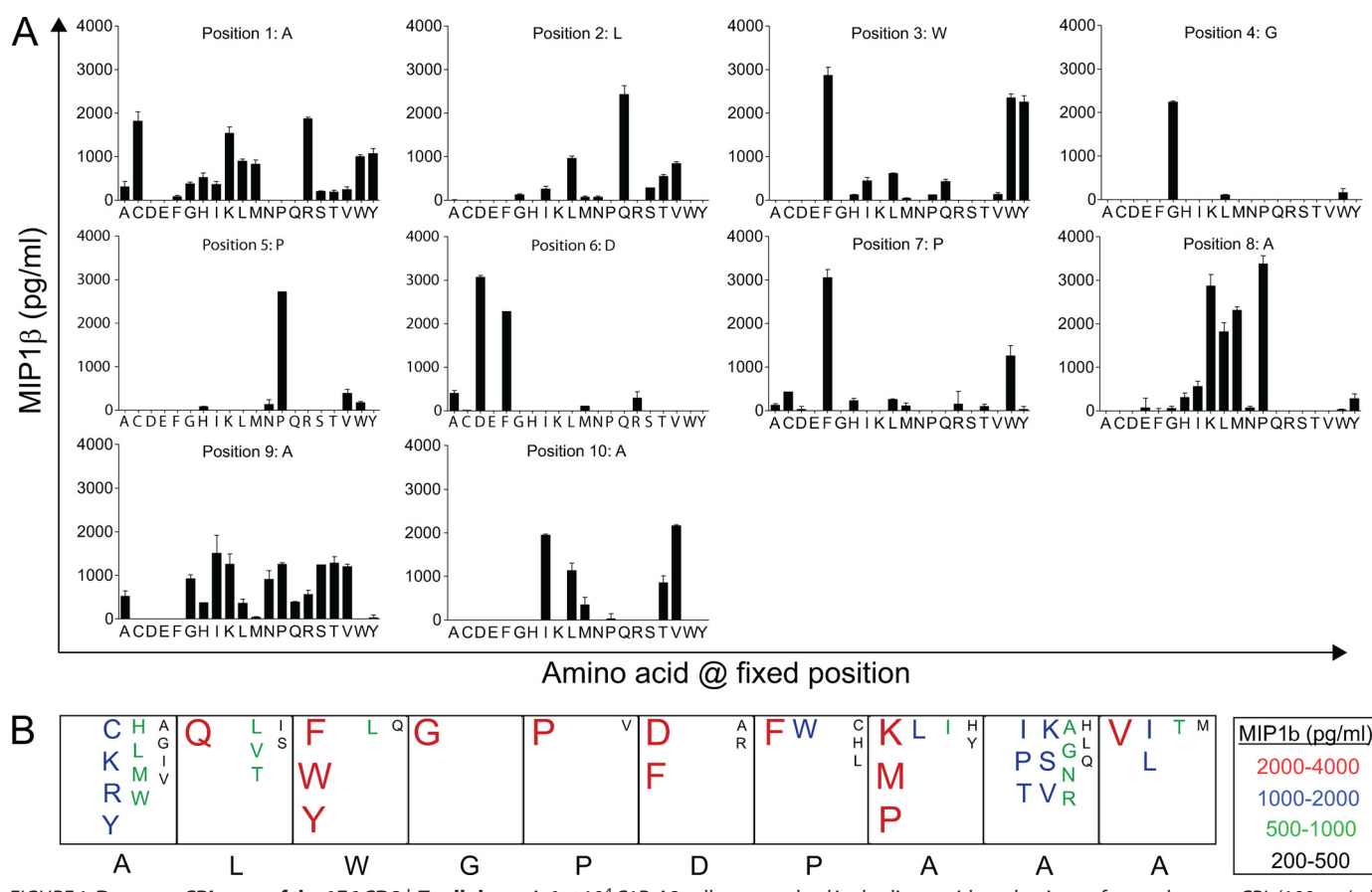


FIGURE 1. **Decamer CPL scan of the 1E6 CD8⁺ T cell clone.** A, 6×10^4 C1R-A2 cells were pulsed in duplicate with each mixture from a decamer CPL (100 μ g/ml) at 37 °C. After 2 h, 3×10^4 1E6 CD8⁺ T cells were added and incubated overnight. Supernatant was harvested and assayed for MIP1 β . B, data from A are displayed as a box plot summary. The index insulin peptide sequence is shown below the boxes in black.

It is unclear how the limited naïve T cell pool responds to a multitude of ligands that it has never encountered before and cannot adapt to at the protein sequence level. One possibility is that T cell immunity inherently features an extremely high level of receptor degeneracy, enabling each TCR to recognize multiple peptides. However, clonal selection theory suggests that individual T cells are specific for a single peptide-MHC molecule with recognition of alternative ligands unlikely. In contrast, studies published in the 1990s demonstrated that T cells can recognize several peptides (5–11). Since then, observations of TCR degeneracy have continued to accumulate in the literature (12–19). In addition, other studies have shown that TCRs can recognize distinct peptides in the context of non-self-MHC, a phenomenon known as alloreactivity (20–25). The majority of previous studies of TCR degeneracy have examined sets of between 2 and 200 peptides, with one recent study examining ~4,000 peptides (19). Given that the entire universe of decamer peptides alone comprises $>10^{13}$ distinct amino acid sequences, the proportion of the peptide universe at this length that has been examined in the most comprehensive study to date (19) remains extremely small ($<0.000000045\%$).

The aim of this study was to probe the entire decamer peptide universe systematically to quantify how many peptides a single TCR can recognize in the context of a single MHC molecule. We demonstrate an unprecedented level of receptor degeneracy that allows a single monoclonal T cell to respond to >one million distinct peptides. As such, the TCR represents

one of the most remarkable biological receptors and by far the most promiscuous known.

EXPERIMENTAL PROCEDURES

Generation and Maintenance of an Autoimmune CD8⁺ T Cell Clone

The 1E6 CD8⁺ T cell clone specific for the human leukocyte antigen (HLA) A*0201-restricted autoantigen preproinsulin peptide ALWGPDPAAA (PPI_{15–24}) was generated as described previously (26).

Decamer Combinatorial Peptide Library (CPL) Scan

The decamer combinatorial peptide library (Pepscan) contains a total of 9.36×10^{12} ($= (10 + 19) \times 19^9$) different decamer peptides and is divided into 200 different peptide mixtures (see Fig. 1). In every peptide mixture, one position has a fixed L-amino acid residue and all other positions are degenerate, with the possibility of any one of 19 natural L-amino acids being incorporated in each individual position (cysteine is excluded). Each library mixture consists of 3.2×10^{11} (19^9) different decamer peptides in approximately equimolar concentrations. For CPL screening, 1E6 CD8⁺ T cells were washed and rested overnight in RPMI 1640 medium containing 100 units/ml penicillin, 100 μ g/ml streptomycin, 2 mM L-glutamine, and 2% heat-inactivated fetal calf serum (all Invitrogen). In 96-well U-bottom plates, 6×10^4 C1R-A2 cells were incubated with various peptide library mixtures (at 100 μ g/ml) in dupli-

The Promiscuity of an Autoimmune TCR

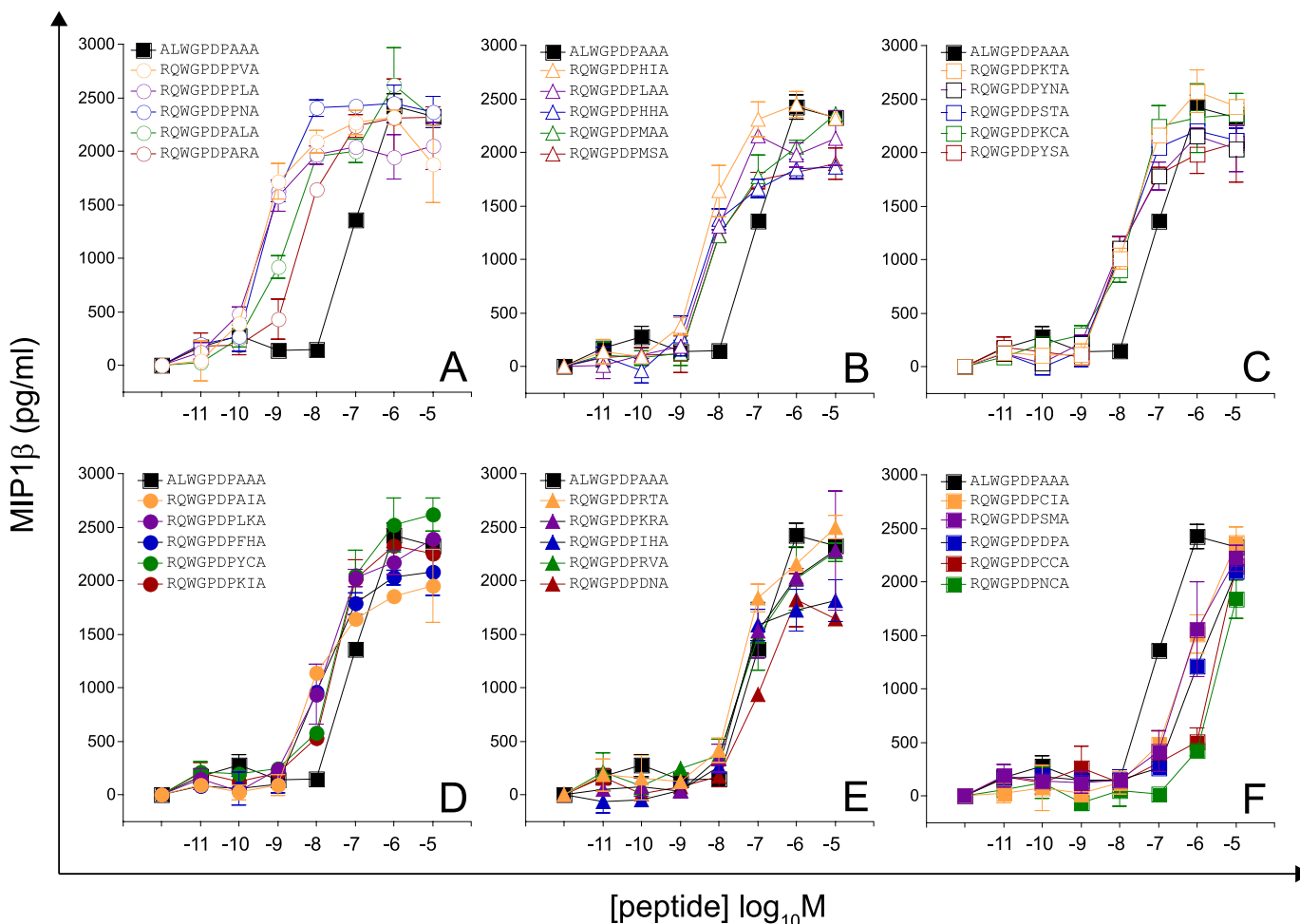


FIGURE 2. Recognition of 30 peptides sampled at random from a large peptide set (motif, RQWGDP{A/C/D/F/H/I/K/L/M/N/P/R/S/V/Y}{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}A; total set size = 225). 6×10^4 C1R-A2 cells were pulsed with peptides at various concentrations. After 2 h, 3×10^4 1E6 CD8⁺ T cells were added and incubated overnight. Supernatant was harvested and assayed for MIP1 β . Each panel displays titrations of five different peptides relative to index. A, titrations of peptides with the highest functional sensitivities. F, titrations of peptides with the lowest functional sensitivities. Error bars, S.D. from the mean of two replicates. pEC_{50} values for each peptide are displayed in supplemental Table S3.

cate for 2 h at 37 °C. Following peptide pulsing, 3×10^4 1E6 CD8⁺ T cells were added, and the assay was incubated overnight at 37 °C. Subsequently, the supernatant was harvested and assayed for MIP1 β by ELISA according to the manufacturer's instructions (R&D Systems).

CD8⁺ T Cell Effector Function Assays: MIP1 β ELISA

Individual peptide sets were assayed for agonist activity by MIP1 β ELISA as described above. Functional sensitivity is expressed by the pEC_{50} of each peptide with respect to the TCR. This is defined as $-1 \times$ the base 10 logarithm (p) of the 50% efficacy concentration (EC_{50}); a greater functional sensitivity is indicated by a larger pEC_{50} value, which was estimated as described in supplemental Equations.

Overview of Sampling Approaches Used to Quantify TCR Degeneracy

Although the TCR has an appreciable degeneracy, it is still specific enough that sampling peptides at random will most likely result in less than ~ 10 strong agonists for every $\sim 10,000$ peptides sampled. For this reason, we employed conditioned sampling. In one approach, we sampled from motif-restricted

peptide sets. This results in lower bound estimates of the actual number of agonist ligands, in that any agonist not fitting the prescribed motif is excluded from the sample. In general, more stringent motifs exclude more agonists but provide better resolution at the high pEC_{50} end of the curve, and vice versa; for this reason, a range of motifs of varying stringency was used. In a second approach (CPL-based importance sampling), we sampled from the entire peptide universe with bias toward peptides that were likely to elicit a response, then estimated a true distribution by applying a correction weighting to the observations (*i.e.* dividing back by the bias).

Sampling Equations

Motif-restricted Sampling—A sampling motif specifies, for each of the m positions in an m -mer peptide, one or more amino acid residues that may occur at that position. If n_p alternative residues have been specified at position p , the probability that a sampled peptide has a given residue at position p equals $1/n_p$ if the given residue is one of the n_p given alternatives, and zero otherwise. Consider a sample of n peptides, consisting of peptides $P^{[1]}$, $P^{[2]}$, ..., $P^{[n]}$, with measured functional sensitivities $pEC_{50}^{[1]}$, $pEC_{50}^{[2]}$, ..., $pEC_{50}^{[n]}$. The cross-reactivity of the TCR is

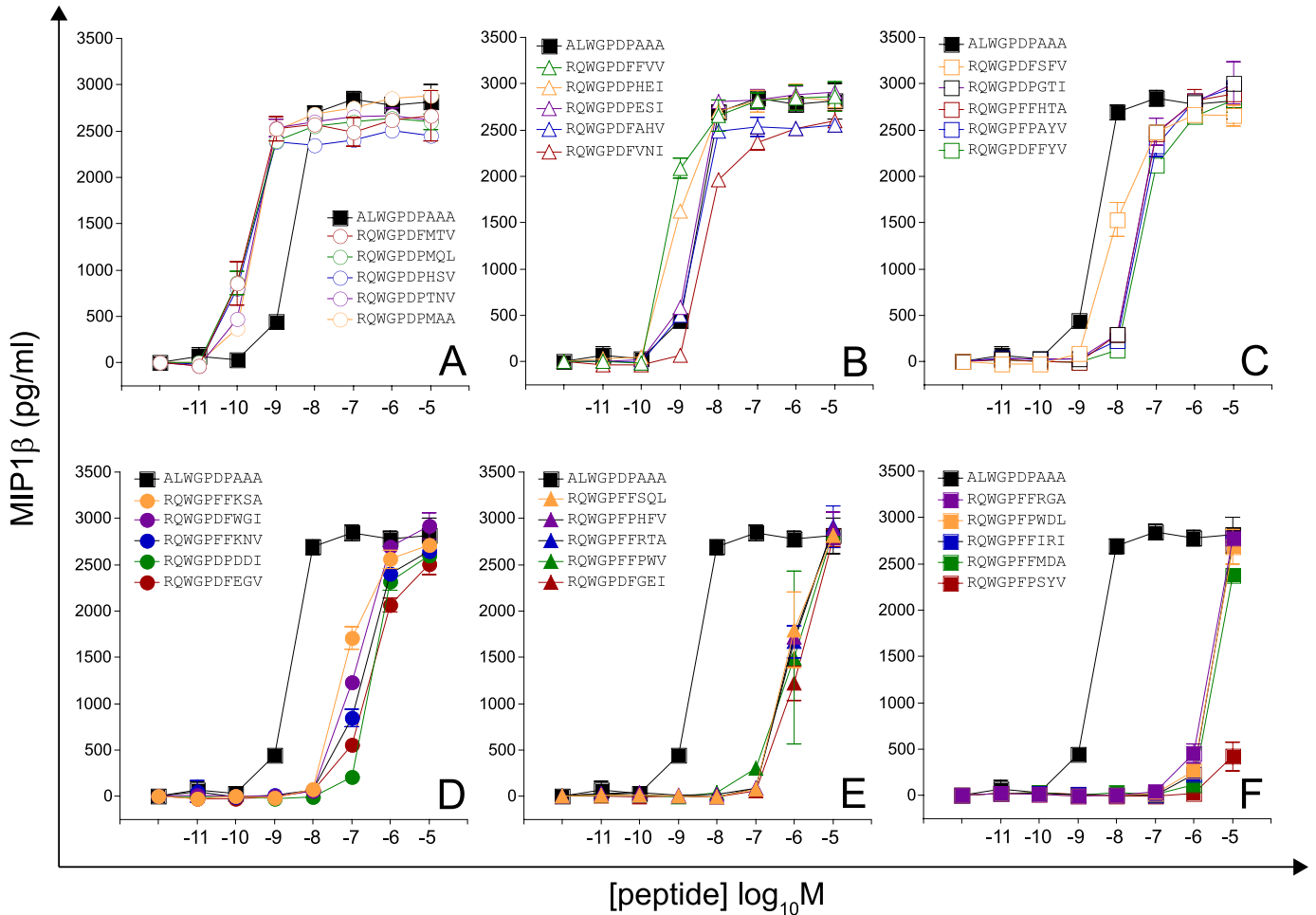


FIGURE 3. Recognition of 30 peptides sampled at random from a large peptide set (motif, RQWGP{D/F}{P/F}XX{A/I/L/V}; total set size = 5,776). Details are as described for Fig. 2.

expressed by the number of peptides whose pEC_{50} exceeds a given value ω . The degeneracy of the TCR is then represented by determining this number for a range of ω values, estimated as follows.

$$N[pEC_{50} > \omega] \geq \frac{\sum_{j=1}^n I(pEC_{50}^{(j)} > \omega)}{n} \prod_{p=1}^m n_p \quad (\text{Eq. 1})$$

Here, $N[pEC_{50} > \omega]$ is the number of peptides that have a pEC_{50} for the TCR that exceeds ω , and

$$I(pEC_{50}^{(j)} > \omega) = 1 \quad (\text{Eq. 2})$$

if sample peptide j has a pEC_{50} that exceeds ω ;

$$I(pEC_{50}^{(j)} > \omega) = 0 \quad (\text{Eq. 3})$$

otherwise. The sampling motif typically excludes a number of agonists (unless $n_p = 20$ at every position, *i.e.* the “universal motif”, which was not used). This exclusion means that the quantity on the right in Equation 1 always underestimates the true $N[pEC_{50} > \omega]$. This means that the motif-based method provides a lower boundary to TCR degeneracy, *i.e.* a conservative estimate.

CPL-based Importance Sampling—The idea of sampling motifs can be generalized to the well known strategy of importance sampling by specifying, for all positions and amino acids,

the probability that a given amino acid occurs at a given position. Then, the probability of drawing a peptide P that has amino acid residue $P(p)$ at position p is given by

$$\mathbb{P}[P] = \mathbb{P}[P(1)P(2)\cdots P(m)] = \prod_{p=1}^m \mathbb{P}[P(p)] \quad (\text{Eq. 4})$$

where $\mathbb{P}[P(p)]$ is the probability of drawing amino acid P at position p . Standard rules of probability stipulate that for each fixed p , the $\mathbb{P}[P(p)]$ sum to unity over the 20 amino acids. The CPL was used to generate distributions that could be expected to bias the sample toward good agonists. However, to correct for the bias in the estimate, observations must be weighted by the reciprocal of the bias. Accordingly, degeneracy was estimated as follows:

$$N[pEC_{50} > \omega] \approx n \exp[H] \frac{\sum_{j=1}^n \mathbb{P}[p^{(j)}]^{-1} I(pEC_{50}^{(j)} > \omega)}{\sum_{j=1}^n \mathbb{P}[p^{(j)}]^{-1}} \quad (\text{Eq. 5})$$

where H is the entropy of the sampling distribution, defined by

$$H = -\sum_{p=1}^m \sum_{P=1}^{20} \mathbb{P}[P(p)] \ln \mathbb{P}[P(p)]. \quad (\text{Eq. 6})$$

The Promiscuity of an Autoimmune TCR

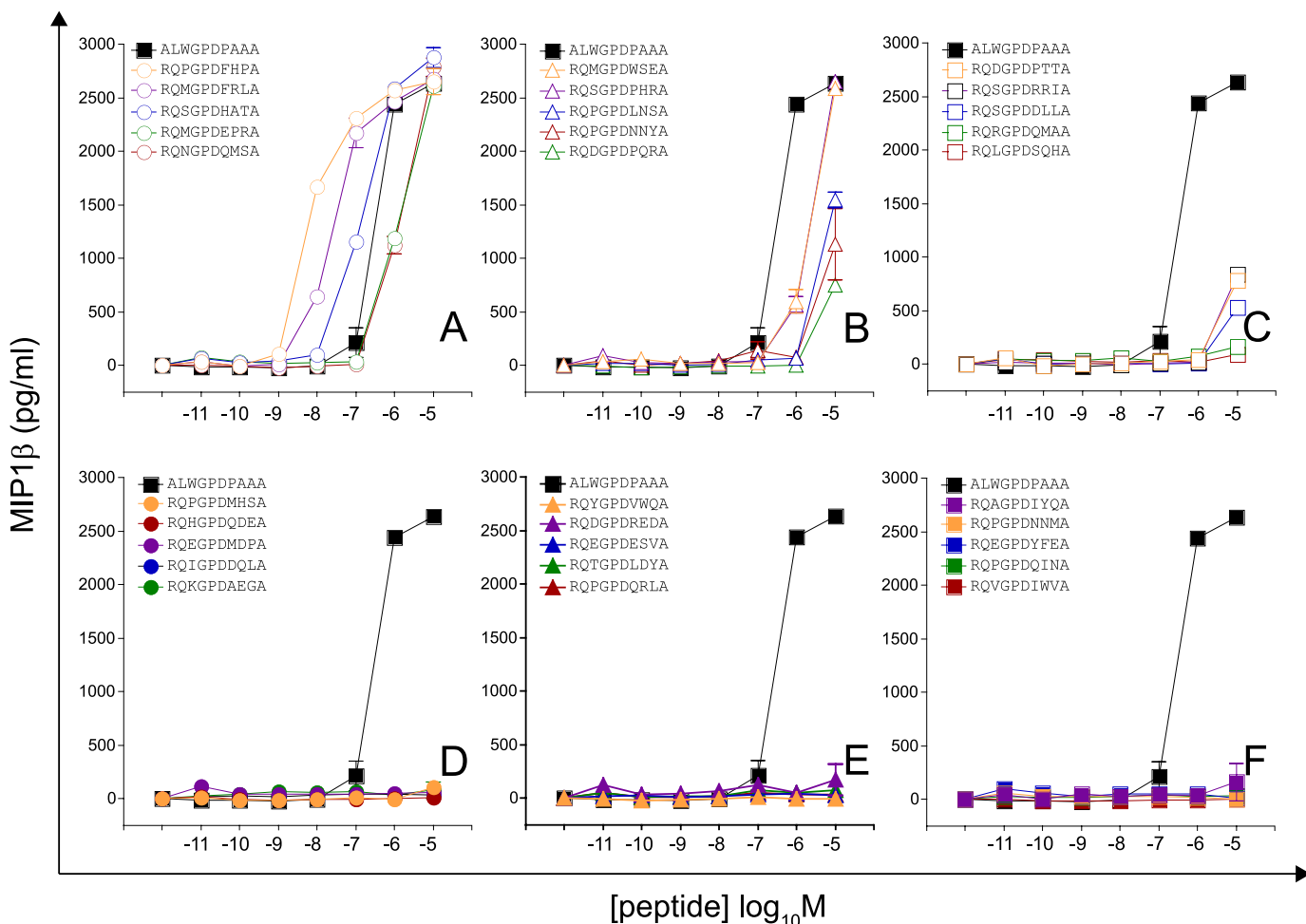


FIGURE 4. Recognition of 30 peptides sampled at random from a large peptide set (motif, RQXGPDXXXA; total set size = 19⁴). Details are as described for Fig. 2.

The factor $n \exp\{H\}$ in Equation 5 is an estimate of the effective sample size. This estimate is based on a standard result concerning the probability of sampling a fixed set of n items. A well known theorem (27) states that this probability is close to $\exp\{-nH\}$. On the other hand, in unbiased sampling from a set of size N , the probability of obtaining any given n -element sample is exactly equal to N^{-n} . Together, these observations indicate that $\exp\{H\}$ can be interpreted as the “diameter” of the population from which the peptide is drawn. Consequently, an n -element sample of m -mer peptides probes an effective set size of

$$20^m(1 - (1 - \exp\{H\}20^{-m})^n) \approx n \exp\{H\}; \quad (\text{Eq. 7})$$

the approximation is accurate when n is tiny compared with $\exp\{-H\}$, as is the case in the experiments reported here.

RESULTS

CPL Screening Reveals the Potential for TCR Degeneracy—To determine the extent of T cell cross-reactivity, we probed the peptide recognition degeneracy of the autoimmune CD8⁺ T cell clone 1E6 using a CPL comprising 9.36×10^{12} different decamer peptides (Fig. 1). Using this approach, we were able to scan every amino acid at every position of the peptide within a random residue “backbone” and build a detailed picture of the

molecular landscape preferred by the 1E6 TCR. The 1E6 clone was generated from a patient with type 1 diabetes and is the only documented example of an autoreactive CD8⁺ T cell that can kill human pancreatic islet β -cells (26). Killing is mediated via recognition of residues 15–24 (ALWGPDPAAA) of the autoantigen preproinsulin bound to HLA A*0201 on the β -cell surface. The majority of HLA A*0201⁺ patients with type I diabetes recognize the preproinsulin 15–24 epitope (26), and HLA A*0201 is known to confer an increased risk of disease (28, 29). The number of amino acids that were recognized by the 1E6 clone was restricted in the central region of the peptide (residues 4–6), suggesting that this TCR makes the majority of its peptide contacts with these residues. In contrast, recognition was highly degenerate at the remaining positions. This degeneracy was confirmed by the ability of 1E6 T cells to recognize a panel of peptides robustly with any of the 20 natural proteogenic L-amino acids at peptide position 8 (supplemental Fig. S1), with half of the substitutions leading to increased levels of functional sensitivity. Similar results were obtained with corresponding scans at other degenerate positions (data not shown). The CPL scan results also revealed that the index peptide is suboptimal in all positions outside the central region (residues 4–6). Thus, positional peptide degeneracy is extreme at 7 of 10 positions, hinting at the

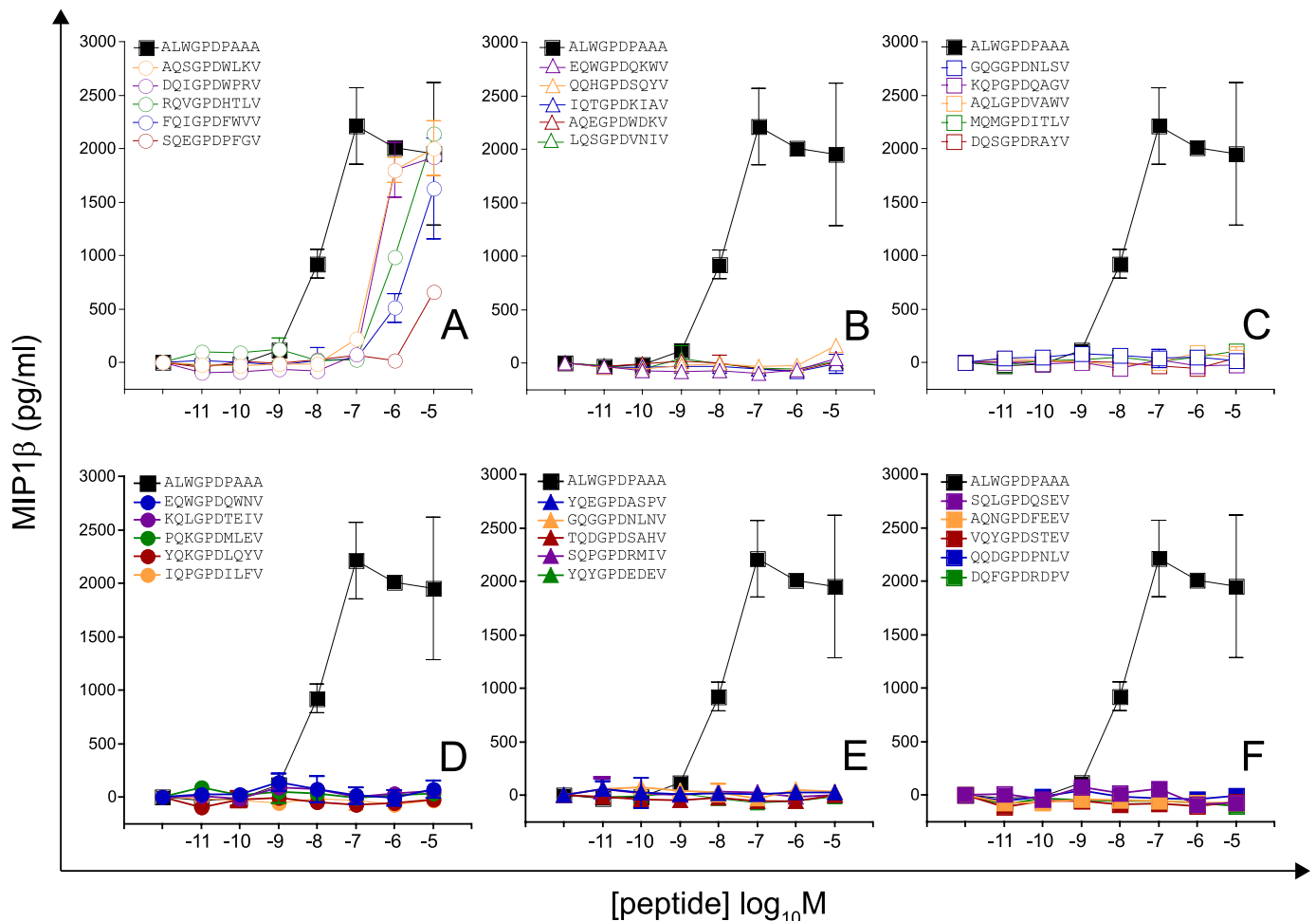


FIGURE 5. Recognition of 30 peptides sampled at random from a large peptide set (motif, XQXGPDXXXV; total set size = 19⁵). Details are as described for Fig. 2.

potential for a single TCR to recognize a multitude of different amino acid combinations.

Quantifying the Number of Decamer Peptides That Can Be Recognized by a Single TCR—We next sought to quantify the number of decamer peptides that can be recognized by the 1E6 clone. The total number of decamers that can be made by combinations of the 20 natural proteogenic L-amino acids is 20¹⁰ (1.02 × 10¹³). Only a small proportion (~1–3%) of all peptides are predicted to bind any given MHC (3, 19), although our own experiments predict that this percentage could be far greater for HLA A*0201 (data not shown). Even with the most conservative estimates of MHC binding (19), the number of potential antigenic HLA A*0201-restricted decamer peptides is still extremely large (1.02 × 10¹¹) and precludes screening all possibilities in T cell recognition assays. To overcome this problem, we screened sets of 30 peptides sampled (Mathematica®; supplemental Fig. S5) from larger motif-restricted or CPL-based importance-sampled sets differing in total size from 225 to 1.66 × 10⁸ individual peptides as described below.

Quantifying T Cell Cross-reactivity Using Motif-restricted Sampling—Motif-restricted peptide sets were designed based on CPL evidence for amino acid preference at each position of the peptide. First, we screened 30 peptides sampled at random

from a total set size of 225 (Motif I, RQWGPDP{A/C/D/F/H/I/K/L/M/N/P/R/S/V/Y}{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}A; Fig. 2 and supplemental Fig. S2). Values of pEC_{50} (−1 × the base 10 logarithm of the EC_{50}) as a measure of functional sensitivity were estimated for all peptides using simultaneous curve fitting (as described in supplemental Equation S1, Fig. S6, and Table S3). Accordingly, increases in functional sensitivity translate into increases in the pEC_{50} value. The 1E6 clone recognized all peptides within this subset efficiently, with 24 of 30 peptides eliciting greater levels of functional sensitivity than the index peptide. A further 30 peptides were sampled at random from a total set size of 5,776 (Motif II, RQWGP{D/F}{P/F}XX{A/I/L/V}, where X denotes any one of the amino acids excluding cysteine (Fig. 3 and supplemental Fig. S2 and Table S3). One peptide from this subset was recognized poorly; 16 were recognized with $pEC_{50} > 7$, with a total of 8 peptides recognized more efficiently than the index peptide (Fig. 3 and supplemental Fig. S2). A further two motif-restricted sets of increasing degeneracy were screened (Motif III, RQXGPDXXXA, total set size 19⁴; and Motif IV, XQXGPDXXXV, total set size 19⁵; X denotes any one of the amino acids excluding cysteine; Figs. 4 and 5 and supplemental Fig. S3 and Table S3). The extent of peptide recognition when peptides are sampled at random from large subsets demonstrates the

The Promiscuity of an Autoimmune TCR

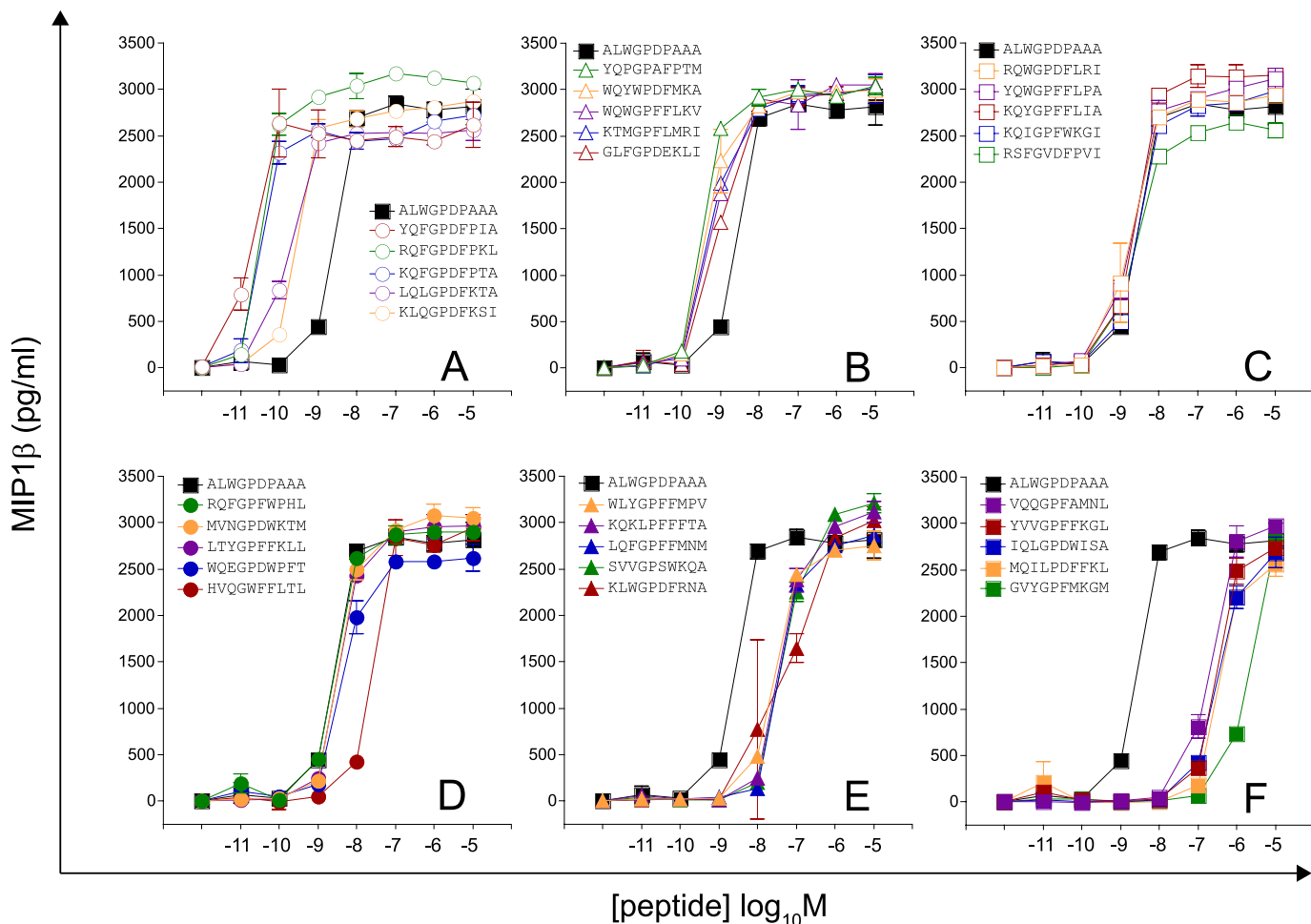


FIGURE 6. Recognition of 30 peptides drawn from a CPL-based importance sampling set with effective size = 1.66×10^8 (calculated from the sampling entropy) (first set). Details are as described for Fig. 2.

considerable degree of degeneracy exhibited by the clonal I E6 TCR.

Quantifying T Cell Cross-reactivity Using CPL-based Importance Sampling—A limitation of sampling from motif-restricted sets is that strong ligands will invariably be excluded, resulting in an underestimate of the true number of agonists (*i.e.* a lower bound estimate of the number of different peptides that a single TCR can recognize). To obtain a more accurate estimate of true T cell cross-reactivity, we employed CPL-based importance sampling, which makes no assumptions about TCR contact or MHC class I-binding residues. Importance sampling ensures that every peptide has a chance of being sampled (although cysteine was excluded to avoid the potential for oxidation), but incorporates bias toward strong agonists predicted using CPL scan data. This bias is adjusted for to yield unbiased estimates of agonist numbers (see under “Sampling Equations”). Raw data from the primary CPL scan were modified as described in [supplemental Table S1](#) and subsequently normalized ([supplemental Table S2](#)) to provide a peptide sampling distribution biasing the sample toward good agonists. The chance of picking a peptide is the product of the normalized weights assigned to each amino acid residue at each given peptide position. Two sets of 30 peptides were drawn from an effective set size of 1.66×10^8 (calculated from the sampling

see under “Sampling Equations,” Equation 7). Of a total of 60 peptides, 34 were recognized efficiently with a $pEC_{50} > 7$, and 22 were better agonists than the index peptide (Figs. 6 and 7 and [supplemental Fig. S4 and Table S3](#)). Interestingly, just this set of 60 peptides identified four peptides with functional sensitivities 100-fold better than the index peptide, despite differing from the index peptide sequence (ALWGPDPAAA) at 6 (YQFGPDFPIA, KQFGPDFPTA) or 7 (RQFGPDFPKL, RQFGPDFPTI) positions (Figs. 6 and 7 and [supplemental Fig. S4 and Table S3](#)). Thus, CPL-based importance sampling demonstrates that a large proportion of peptides from a biased set of 1.66×10^8 peptides would be recognized and that these recognized peptides can differ considerably from the index peptide.

A Single TCR Can Recognize More Than One Million Different Peptides—The pEC_{50} expresses the potency of a ligand, often referred to as “functional avidity” ([supplemental Fig. S6](#)). Cross-reactivity can be quantified precisely by specifying the number of ligands that the TCR recognizes with a pEC_{50} of at least a given value ω . This number decreases as ω increases; an insight into the nature of TCR degeneracy is afforded by plotting this agonist number as a function of ω . Estimation of this number was performed using Equations 1 and 5, resulting in Fig. 8. The motif-restricted estimate (*solid lines* in Fig. 8) is a lower boundary, which becomes tighter as the degeneracy of the motif

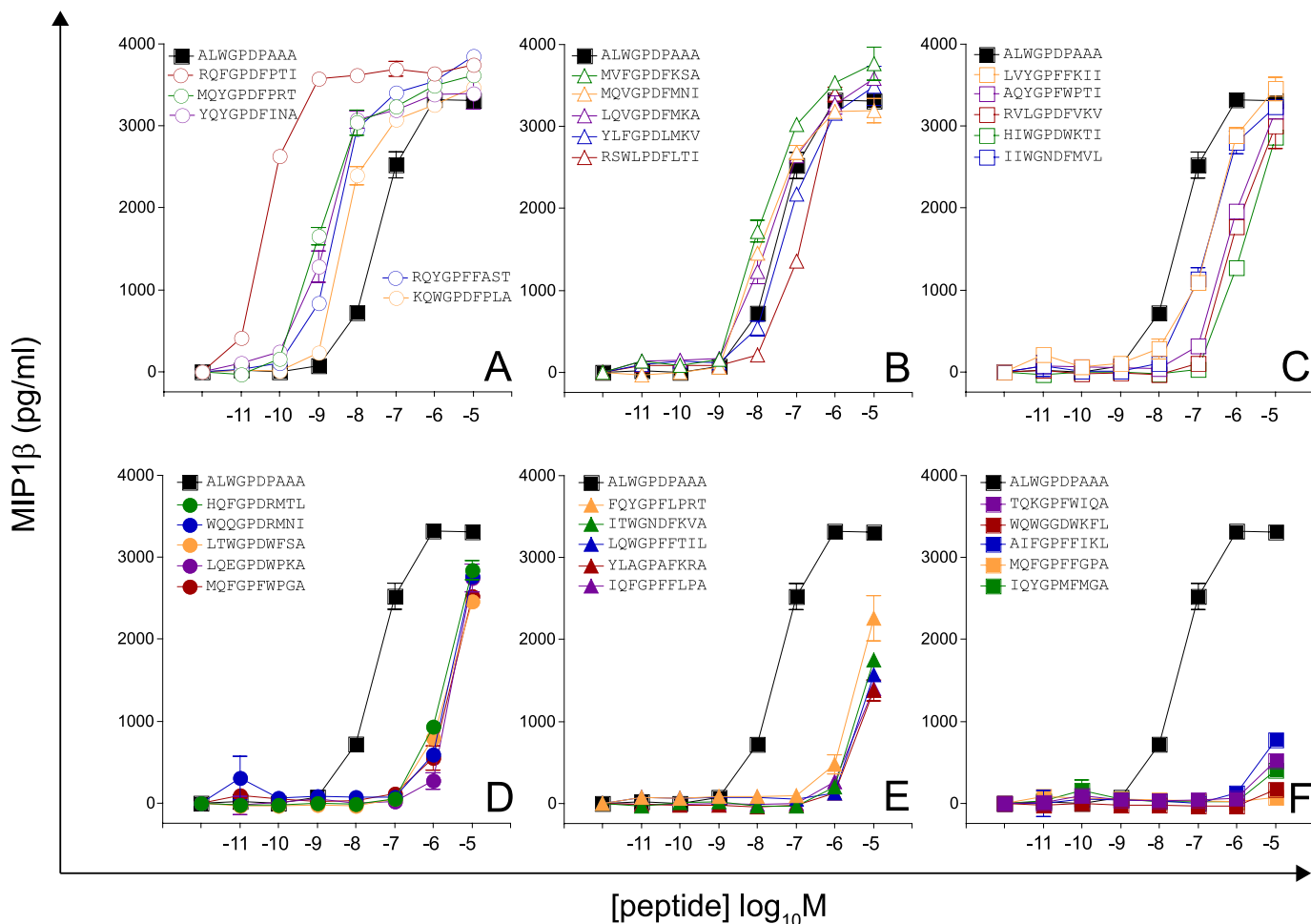


FIGURE 7. Recognition of 30 peptides drawn from a CPL-based importance sampling set with effective size = 1.66×10^8 (calculated from the sampling entropy) (second set). Details are as described for Fig. 2.

increases; however, this advantage is offset by the decreasing chances of finding good agonists in a sample of 30 peptides. The estimates based on the biased sampling (*dashed lines* in Fig. 8) indicate that in the order of one million agonists exist for 1E6 that are as least as good as the index peptide. For comparison, the curve derived from TCR activation theory (30, 31) is also shown (*gray dashed line* in Fig. 8; [supplemental Equation S2](#)).

DISCUSSION

Despite the huge potential importance of TCR degeneracy to human health, there has never been a comprehensive attempt to quantify the number of peptides that can be recognized by a single TCR. To address this issue, we examined the extent of cross-reactivity exhibited by a single autoimmune T cell clone with pathogenic relevance in human type I diabetes (1E6). Our analysis demonstrates that the 1E6 TCR can recognize ~ 500 peptides within a factor 2 of the optimal agonist (*i.e.* peptides that have a functional sensitivity that is at least 50% of the functional sensitivity of the optimal agonist). An estimated 60,000 peptides have a functional sensitivity within a factor 10 of the optimal agonist, and $\sim 1.3 \times 10^6$ peptides are within a factor 100 of the optimal agonist. These considerations are especially significant given that the functional sensitivity of 1E6 CD8⁺ T cells for the index peptide is at least 100-fold lower than the optimal agonist; this is illustrated by RQFGPDFPTI, the functional sensitivity of which

is 100-fold better than the index. Almost 10 million peptides are within a factor 1,000 of the optimal agonist, but such weak agonists will not generally be physiologically significant unless presented at very high copy numbers. Taken together, these results indicate that the 1E6 TCR has >one million significant peptide agonists at concentrations with the potential to be physiologically relevant.

When putting TCR degeneracy into perspective, it is important to realize that individual TCRs capable of recognizing 10^6 decamer peptides still only have a less than 1 in 10^7 chance of cross-reacting with any peptide chosen at random from the entire decamer peptide universe ($\sim 10^{13}$). A high level of cross-reactivity is therefore amply compatible with the degree of specificity required for self/non-self determination. Furthermore, the number of decamer peptides that it is possible to make from the entire human proteome (excluding post-translational amino acid modifications) is only one millionth of the possible peptide universe at this length. Functional recognition of $\sim 10^6$ decamer peptides by a single TCR translates into a frequency of cross-reactivity of 1:100,000 (assuming that 1% of peptides bind to MHC), which is likely to be the most accurate estimate of this parameter to date due to the comprehensive nature of our approach. The probability of cross-reactivity with any individual peptide is an important consideration in terms of viral escape, bystander activation and autoimmune side effects,

The Promiscuity of an Autoimmune TCR

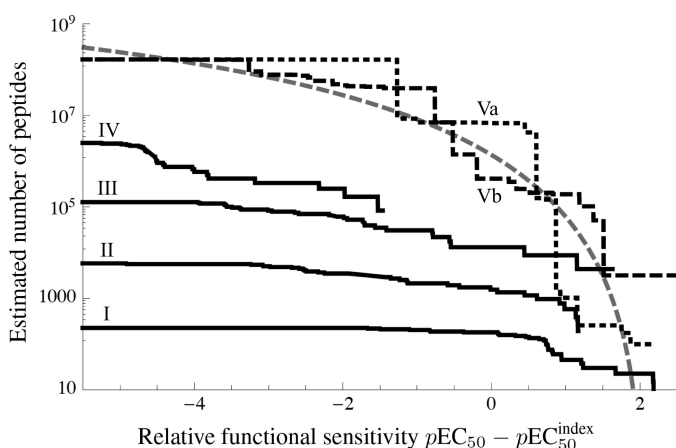


FIGURE 8. 1E6 CD8⁺ T cells can recognize more than one million different decamer peptides. *I*, RQWGPDP{A/C/D/F/H/I/K/L/M/N/P/R/S/N/Y}{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}A (set size 225; 30 peptides sampled at random). *II*, RQWGP{D/F}{P/F}XX{A/I/L/V} (set size 5,776; 30 peptides sampled at random). *III*, RQXGPDXXXA (set size 19⁴; 30 peptides sampled at random). *IV*, XQXGP-DXXXV (set size 19⁵; 30 peptides sampled at random). In the motifs, X denotes any one of the 19 amino acids excluding cysteine. *Va* and *Vb*, two replicates of a biased sampling set (effective set size 1.66×10^8 , calculated from the sampling entropy); each set of 30 peptides was sampled with bias toward strong agonists, where the bias weights were based on the primary CPL scan. Relative functional sensitivities ($pEC_{50} - pEC_{50}^{\text{index}}$) are plotted as survivor curves. *Gray dashed line*, theoretical curve (supplemental Equation S2, with $\alpha = 4.5$, $\beta = 10$, $\gamma = 2$, $N_0 = 20^8$ anchorable peptides). The biased samples (*black dashed and dotted lines*) estimate the TCR degeneracy spectrum, whereas the motif-based samples (*black solid lines*) provide a lower boundary to the TCR degeneracy spectrum.

and the results presented here fit well with theoretical considerations of T cell immunity (3). It should be noted, however, that the degeneracy curves are estimates based on samples that constitute a small fraction of the number of possible peptide ligands and, hence, should be considered as depicting an order-of-magnitude estimate that is supported by the lower bounds inferred from motif-based samples.

The 1E6 CD8⁺ T cell clone was chosen for these studies to highlight the huge potential for T cell cross-reactivity as a possible cause of autoimmunity. In support of the generality of our findings, we have also observed high levels of degeneracy at some positions in a 9-amino acid residue nonautoimmune epitope (32). Furthermore, the recognition of longer, MHC class II-restricted peptides by CD4⁺ T cells with a “TCR footprint” of similar size could ensure that these cells incorporate the capacity to recognize tens, or possibly even hundreds, of millions of peptides at physiologically relevant surface densities (33). The reality of T cell cross-reactivity, as quantified here, has far reaching implications. It provides an explanation for how a limited pool of TCRs can provide the broad antigenic coverage that is required for effective immunity. In addition, the extent of TCR degeneracy suggests that almost any peptide antigen can be improved for any given cognate TCR, in the sense of there being at least one stronger agonist than the original index peptide, thereby providing scope for rational therapeutic interventions based on the directed manipulation of T cell immunity.

Acknowledgments—We thank Don Mason and Bent Jakobsen for critical reading of the manuscript and David Rand for careful scrutiny of the mathematics.

REFERENCES

- Rudolph, M. G., and Wilson, I. A. (2002) The specificity of TCR/pMHC interaction. *Curr. Opin. Immunol.* **14**, 52–65
- Rudolph, M. G., Stanfield, R. L., and Wilson, I. A. (2006) How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466
- Mason, D. (1998) A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19**, 395–404
- Arstila, T. P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999) A direct estimate of the human $\alpha\beta$ T cell receptor diversity. *Science* **286**, 958–961
- Wraith, D. C., Bruun, B., and Fairchild, P. J. (1992) Cross-reactive antigen recognition by an encephalitogenic T cell receptor. Implications for T cell biology and autoimmunity. *J. Immunol.* **149**, 3765–3770
- Bhardwaj, V., Kumar, V., Geysen, H. M., and Sercarz, E. E. (1993) Degrade recognition of a dissimilar antigenic peptide by myelin basic protein-reactive T cells. Implications for thymic education and autoimmunity. *J. Immunol.* **151**, 5000–5010
- Reay, P. A., Kantor, R. M., and Davis, M. M. (1994) Use of global amino acid replacements to define the requirements for MHC binding and T cell recognition of moth cytochrome *c* (93–103). *J. Immunol.* **152**, 3946–3957
- Evavold, B. D., Sloan-Lancaster, J., Wilson, K. J., Rothbard, J. B., and Allen, P. M. (1995) Specific T cell recognition of minimally homologous peptides: evidence for multiple endogenous ligands. *Immunity* **2**, 655–663
- Wucherpfennig, K. W., and Strominger, J. L. (1995) Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* **80**, 695–705
- Hemmer, B., Fleckenstein, B. T., Vergelli, M., Jung, G., McFarland, H., Martin, R., and Wiesmüller, K. H. (1997) Identification of high potency microbial and self ligands for a human autoreactive class II-restricted T cell clone. *J. Exp. Med.* **185**, 1651–1659
- Kersh, E. N., Shaw, A. S., and Allen, P. M. (1998) Fidelity of T cell activation through multistep T cell receptor ζ phosphorylation. *Science* **281**, 572–575
- Kissler, S., Anderton, S. M., and Wraith, D. C. (2002) Cross-reactivity and T-cell receptor antagonism of myelin basic protein-reactive T cells is modulated by the activation state of the antigen presenting cell. *J. Autoimmun.* **19**, 183–193
- Nino-Vasquez, J. J., Allicotti, G., Borrás, E., Wilson, D. B., Valmori, D., Simon, R., Martin, R., and Pinilla, C. (2004) A powerful combination: the use of positional scanning libraries and biometrical analysis to identify cross-reactive T cell epitopes. *Mol. Immunol.* **40**, 1063–1074
- Crawford, F., Huseby, E., White, J., Marrack, P., and Kappler, J. W. (2004) Mimotopes for alloreactive and conventional T cells in a peptide-MHC display library. *PLoS Biol.* **2**, E90
- Lee, J. K., Stewart-Jones, G., Dong, T., Harlos, K., Di Gleria, K., Dorrell, L., Douek, D. C., van der Merwe, P. A., Jones, E. Y., and McMichael, A. J. (2004) T cell cross-reactivity and conformational changes during TCR engagement. *J. Exp. Med.* **200**, 1455–1466
- Kan-Mitchell, J., Bajcz, M., Schaubert, K. L., Price, D. A., Brenchley, J. M., Asher, T. E., Douek, D. C., Ng, H. L., Yang, O. O., Rinaldo, C. R., Jr., Benito, J. M., Bisikirska, B., Hegde, R., Marincola, F. M., Boggiano, C., Wilson, D., Abrams, J., Blondelle, S. E., and Wilson, D. B. (2006) Degeneracy and repertoire of the human HIV-1 Gag p17(77–85) CTL response. *J. Immunol.* **176**, 6690–6701
- Wilson, D. B., Wilson, D. H., Schroder, K., Pinilla, C., Blondelle, S., Houghten, R. A., and Garcia, K. C. (2004) Specificity and degeneracy of T cells. *Mol. Immunol.* **40**, 1047–1055
- Dai, S., Huseby, E. S., Rubtsova, K., Scott-Browne, J., Crawford, F., Macdonald, W. A., Marrack, P., and Kappler, J. W. (2008) Crossreactive T cells spotlight the germ line rules for $\alpha\beta$ T cell-receptor interactions with MHC molecules. *Immunity* **28**, 324–334
- Ishizuka, J., Grebe, K., Shenderov, E., Peters, B., Chen, Q., Peng, Y., Wang, L., Dong, T., Pasquetto, V., Oseroff, C., Sidney, J., Hickman, H., Cerundolo, V., Sette, A., Bennink, J. R., McMichael, A., and Yewdell, J. W. (2009) Quantitating T cell cross-reactivity for unrelated peptide antigens. *J. Immunol.* **183**, 4337–4345
- Udaka, K., Tsomides, T. J., and Eisen, H. N. (1992) A naturally occurring

- peptide recognized by alloreactive CD8⁺ cytotoxic T lymphocytes in association with a class I MHC protein. *Cell* **69**, 989–998
21. Reiser, J. B., Darnault, C., Guimezanes, A., Grégoire, C., Mosser, T., Schmitt-Verhulst, A. M., Fontecilla-Camps, J. C., Malissen, B., Housset, D., and Mazza, G. (2000) Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nat. Immunol.* **1**, 291–297
 22. Archbold, J. K., Macdonald, W. A., Miles, J. J., Brennan, R. M., Kjer-Nielsen, L., McCluskey, J., Burrows, S. R., and Rossjohn, J. (2006) Alloreactivity between disparate cognate and allogeneic pMHC-I complexes is the result of highly focused, peptide-dependent structural mimicry. *J. Biol. Chem.* **281**, 34324–34332
 23. Felix, N. J., and Allen, P. M. (2007) Specificity of T-cell alloreactivity. *Nat. Rev. Immunol.* **7**, 942–953
 24. Archbold, J. K., Ely, L. K., Kjer-Nielsen, L., Burrows, S. R., Rossjohn, J., McCluskey, J., and Macdonald, W. A. (2008) T cell allorecognition and MHC restriction: a case of Jekyll and Hyde? *Mol. Immunol.* **45**, 583–598
 25. Macdonald, W. A., Chen, Z., Gras, S., Archbold, J. K., Tynan, F. E., Clements, C. S., Bharadwaj, M., Kjer-Nielsen, L., Saunders, P. M., Wilce, M. C., Crawford, F., Stadinsky, B., Jackson, D., Brooks, A. G., Purcell, A. W., Kappler, J. W., Burrows, S. R., Rossjohn, J., and McCluskey, J. (2009) T cell allorecognition via molecular mimicry. *Immunity* **31**, 897–908
 26. Skowera, A., Ellis, R. J., Varela-Calviño, R., Arif, S., Huang, G. C., Van-Krinks, C., Zaremba, A., Rackham, C., Allen, J. S., Tree, T. I., Zhao, M., Dayan, C. M., Sewell, A. K., Unger, W. W., Drijfhout, J. W., Ossendorp, F., Roep, B. O., and Peakman, M. (2008) CTLs are targeted to kill beta cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. *J. Clin. Invest.* **118**, 3390–3402
 27. Karlin, S., and Taylor, H. M. (1974) *A First Course in Stochastic Processes*, 2nd Ed., pp. 495–502, Academic Press, New York
 28. Nejentsev, S., Howson, J. M., Walker, N. M., Szeszko, J., Field, S. F., Stevens, H. E., Reynolds, P., Hardy, M., King, E., Masters, J., Hulme, J., Maier, L. M., Smyth, D., Bailey, R., Cooper, J. D., Ribas, G., Campbell, R. D., Clayton, D. G., and Todd, J. A. (2007) Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature* **450**, 887–892
 29. Todd, J. A., Walker, N. M., Cooper, J. D., Smyth, D. J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S. F., Payne, F., Lowe, C. E., Szeszko, J. S., Hafler, J. P., Zeitels, L., Yang, J. H., Vella, A., Nutland, S., Stevens, H. E., Schuilenburg, H., Coleman, G., Maisuria, M., Meadows, W., Smink, L. J., Healy, B., Burren, O. S., Lam, A. A., Ovington, N. R., Allen, J., Adlem, E., Leung, H. T., Wallace, C., Howson, J. M., Guja, C., Ionescu-Tirgoviste, C., Simmonds, M. J., Heward, J. M., Gough, S. C., Dunger, D. B., Wicker, L. S., and Clayton, D. G. (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* **39**, 857–864
 30. van den Berg, H. A., Rand, D. A., and Burroughs, N. J. (2001) A reliable and safe T cell repertoire based on low-affinity T cell receptors. *J. Theor. Biol.* **209**, 465–486
 31. van den Berg, H. A., and Rand, D. A. (2007) Quantitative theories of T-cell responsiveness. *Immunol. Rev.* **216**, 81–92
 32. Wooldridge, L., Laugel, B., Ekeruche, J., Clement, M., van den Berg, H. A., Price, D. A., and Sewell, A. K. (2010) CD8 controls T cell cross-reactivity. *J. Immunol.* **185**, 4625–4632
 33. Hiemstra, H. S., Drijfhout, J. W., and Roep, B. O. (2000) Antigen arrays in T cell immunology. *Curr. Opin. Immunol.* **12**, 80–84

Supplemental Data for

**A single autoimmune T-cell receptor
recognizes over a million different peptides**

Linda Wooldridge[#], Julia Ekeruche-Makinde[#], Hugo A. van den Berg[#], Anna Skowera, John J. Miles, Mai Ping Tan, Garry Dolton, Mathew Clement, Sian Llewellyn-Lacey, David A. Price, Mark Peakman and Andrew K. Sewell

File prepared by LW. E-mail: wooldridgel@cardiff.ac.uk

[#]These authors contributed equally to this study.

Supplemental Equations S1 and S2

Supplemental Figures S1 to S6

Supplemental Tables S1 to S3

Supplemental References

SUPPLEMENTAL EQUATIONS

Estimation of pEC_{50}

Functional sensitivity is expressed by the pEC_{50} of that peptide with respect to the TCR. This is defined as minus 1 times the base-10 logarithm of the 50% efficacy concentration. Accordingly, The read-out y (MIP1 β by ELISA) is related to the incubation concentration C , as follows:

$$y = y_{\min} + \frac{y_{\max} - y_{\min}}{1 + 10^{\kappa(pEC_{50} + \log C)}} \quad (S1)$$

where y_{\min} , y_{\max} , and κ are parameters that were estimated using non-linear least squares. Parsimony was achieved through simultaneous fitting; i.e. y_{\min} , y_{\max} , and κ were assumed to have the same value for all peptides (κ is the steepness of the dose-response curve). Eqn (S1) is derived by assuming, first, that each pMHC molecule contributes a signal w_{ij} where i denotes the TCR clonotype and j the pMHC ligand, so that the combined signal generated by Z_j copies of the ligand present in the contact area between the CD8⁺ T-cell and the C1R-A2 B-cell is given by $Z_j w_{ij}$; the quantity w_{ij} represents the functional sensitivity. Second, it is assumed that the CD8⁺ T-cell is activated when the signal $Z_j w_{ij}$ exceeds a signalling threshold $W_{act}(1)$, which is assumed to follow a log-logistic distribution. Third, it is assumed that Z_j is proportional to the pulsing concentration (the proportionality is unknown but is eliminated by studying the ratio relative to the EC_{50} of the index peptide; i.e. the difference in pEC_{50}). Then, the EC_{50} is inversely proportional to w_{ij} and eqn (S1)

follows by assuming that the fraction of responding CD8⁺ T-cells is proportional to the response above baseline (y_{\min}).

Theoretical curve

For comparison, the curve derived from TCR activation theory (1,2) is also shown (grey dashed line in Figure 4), based on the formula for the number of peptides with a relative functional sensitivity that is at least as strong as ω :

$$\mathcal{N}[pEC_{50} - pEC_{50}^{\text{index}} > \omega] = \frac{N_0}{2} \left(\text{erf} \left[\frac{\beta}{\alpha\sqrt{2}} \left(1 + \alpha - (\omega - \gamma) \log 10 + \mathcal{W}_0 \left(-e^{-1+(\omega-\gamma) \log 10} \right) \right) \right] - \text{erf} \left[\frac{\beta}{\alpha\sqrt{2}} \left(1 + \alpha - (\omega - \gamma) \log 10 + \mathcal{W}_{-1} \left(-e^{-1+(\omega-\gamma) \log 10} \right) \right) \right] \right) \quad (\text{S2})$$

where erf denotes the error function, \mathcal{W} denotes the Lambert W-function, N_0 is the number of MHC-anchorable peptides, α is a location parameter, β is a specificity parameter, and γ is an offset to account for the fact that the index peptide, rather than the absolutely optimal peptide, is used as reference.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1: Position degeneracy of the 1E6 TCR. MIP1 β activation data for a set of peptides with the sequence ALWGPDPxAA, where x is 1 of the 20 natural proteogenic L-amino acids. 6×10^4 C1R-A2 B-cells were pulsed with peptides at various concentrations. After 2 hours, 3×10^4 1E6 CD8⁺ T-cells were added and incubated overnight. Supernatant was harvested and assayed for MIP1 β .

Figure S2: Recognition of peptides sampled at random from large fixed motif sets. A&B. The response of 1E6 CD8⁺ T-cells to 30 peptides sampled at random from a large motif-restricted set:

RQWGPDP{A/C/D/F/H/I/K/L/M/N/P/R/S/V/Y}{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}

A;

total set size = 225). Assays as in Figure S1. **A.** Selected titration curves are shown to demonstrate the range of functional sensitivities observed within the set of 30 peptides. Standard deviation from the mean of two replicates is shown. **B.** The functional sensitivities of all peptides tested are displayed relative to that of the index peptide ($pEC_{50} - pEC_{50}$ index) to control for variations in absolute values between assays. Peptides with a functional sensitivity equivalent to index have a value 0, peptides with a functional sensitivity greater than index have values >0 and peptides that are less immunogenic than index have values <0 . Coloured bars match the key shown in panel A for individual peptides in the set. **C&D.** Details as for A&B, except that the motif is RQWGP{D/F}{P/F}xx{A/I/L/V} where x denotes any one of the 19 amino acids excluding cysteine and the total set size = 5776.

Figure S3: Recognition of 30 peptides sampled at random from large peptide sets (motif: RQxGPDxxxA; total set size = 19^4 or xQxGPDxxxV; total set size = 19^5).

Assays as in Figure S1. The functional sensitivities of all peptides tested are displayed relative to that of the index peptide ($pEC_{50} - pEC_{50}$ index).

Figure S4: Recognition of peptides selected by CPL-based importance sampling.

A&B. $1E6$ $CD8^+$ T-cell recognition of two sets of 30 peptides sampled from a CPL-based importance sampling set with effective size = 1.66×10^8 (calculated from the sampling entropy). Assays as described in Figure S1 legend.

Figure S5: *Mathematica* code. Peptides were sampled at random from larger motif-restricted or CPL-based importance sampling sets varying in total size from 225 to 1.66×10^8 individual peptides using Mathematica[®] (Wolfram Research Europe Ltd., Long Hanborough, UK). Displayed are the codes to generate the peptide samples as well as an example of the simultaneous non-linear curve fitting procedure. The workspace file is available upon request.

Figure S6: The functional sensitivity of $1E6$ $CD8^+$ T-cells to all peptide ligands tested.

Simultaneous curve fitting (as described in supplementary equations; eqn S1) was used to estimate functional sensitivity measured as pEC_{50} for peptides sampled from:

A:

RQWGPDP{A/C/D/F/H/I/K/L/M/N/P/R/S/V/Y}{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}

A (set size 225; 30 peptides sampled at random); **B:**

RQWGP{D/F}{P/F}xx{A/I/L/V} (set size 5776; 30 peptides sampled at random); **C:**

RQxGPDxxxA (set size 19^4 ; 30 peptides sampled at random); **D**: xQxGPDxxxV (set size 19^5 ; 30 peptides sampled at random) and **E&F**: two replicates of a biased sampling set (effective set size 1.66×10^8 , calculated from the sampling entropy); each set of 30 peptides was sampled with bias towards strong agonists, weighted based on the primary CPL scan. Values for each peptide are displayed in Table S3.

SUPPLEMENTAL TABLE LEGENDS

Table S1: Data matrix used for derivation of biased sampling set. First, raw data values generated from the decamer CPL scan (MIP1 β levels in pg/ml) were inserted into the table. The following modifications were then made, based on results obtained with previous peptide screening of the 1E6 clone (data not shown): cysteine set to zero; small or negative values set to 5; position 2: double weight Q; position 3: assign 400 to A, E, K and N, turn V up to 400; position 4: tune L and W down to 50; position 5: tune all responses down to 50 except for P; position 6: tune all responses down to 50 except R & M; position 7: leave as original screen data; position 8: 500 for A, 200 for F, R and V; position 9: 200 for F; position 10: tune V down to 1000, add value for index peptide for A and increase weight of L to 2000.

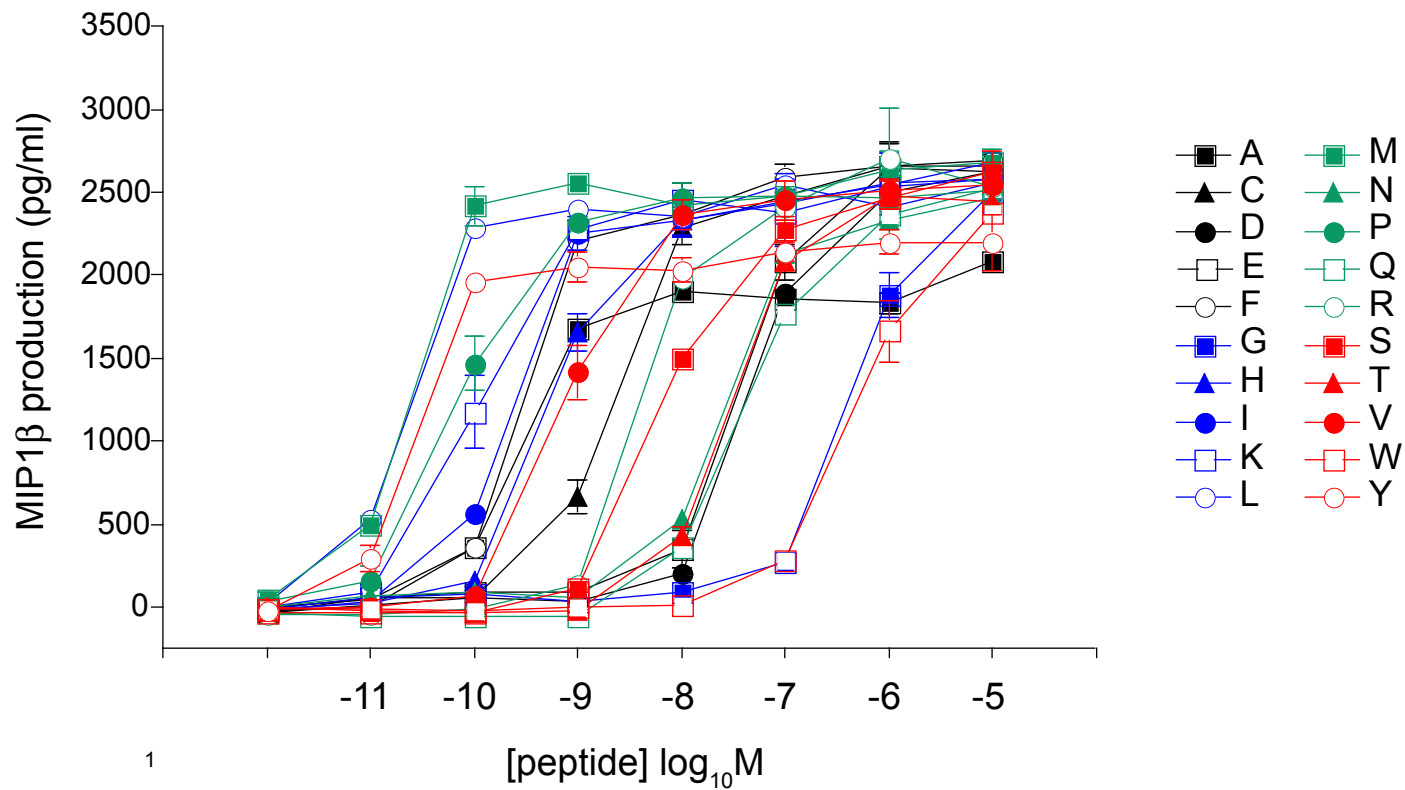
Table S2: Normalized matrix used for derivation of biased sampling set. All MIP1 β readings from Table S1 were given a value between 0-1.

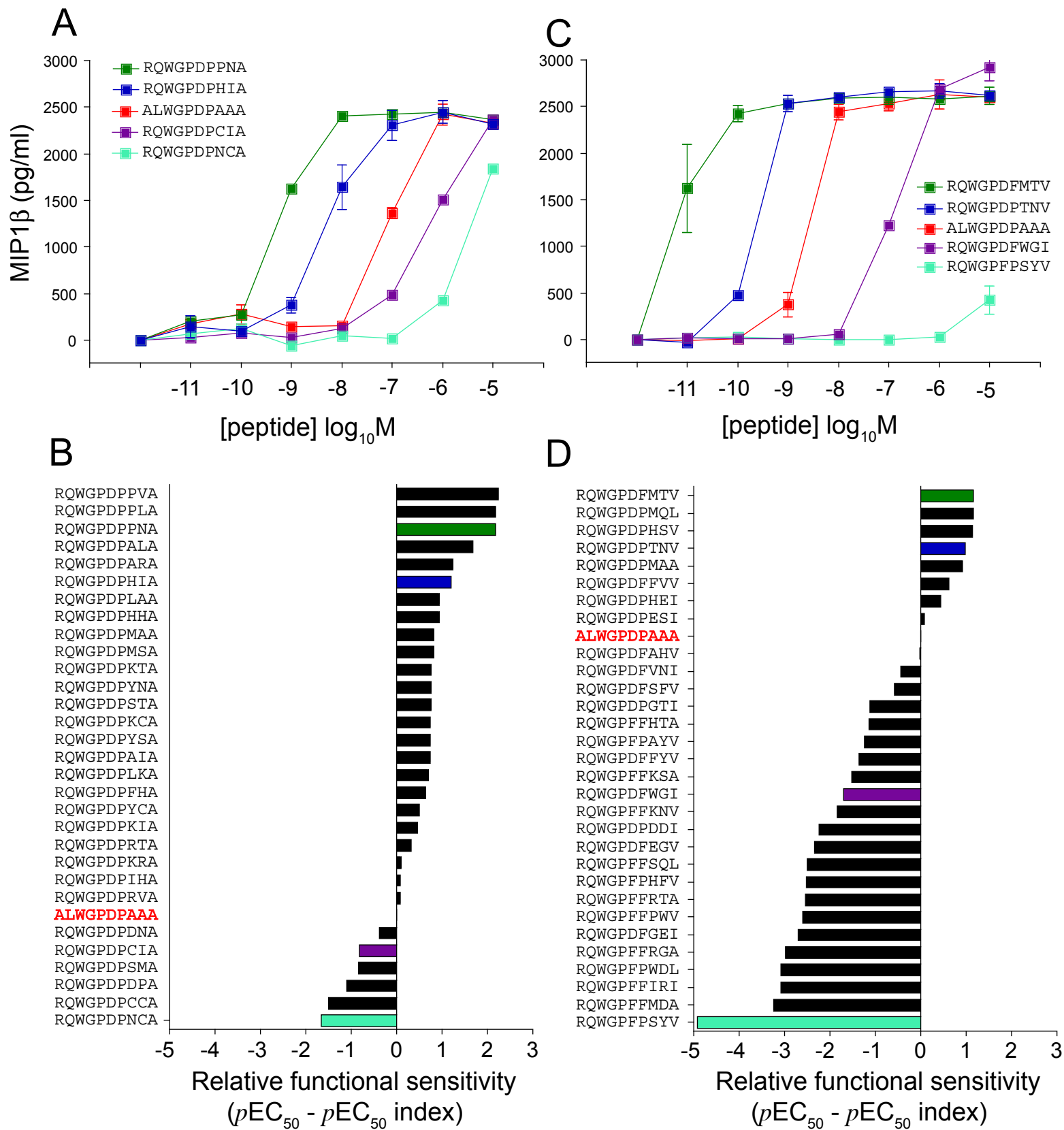
Table S3: pEC_{50} values for all peptide ligands tested. Simultaneous curve fitting (as described in supplementary equations) was used to estimate functional sensitivity measured as pEC_{50} for every peptide tested (see Figure S6).

SUPPLEMENTAL REFERENCES

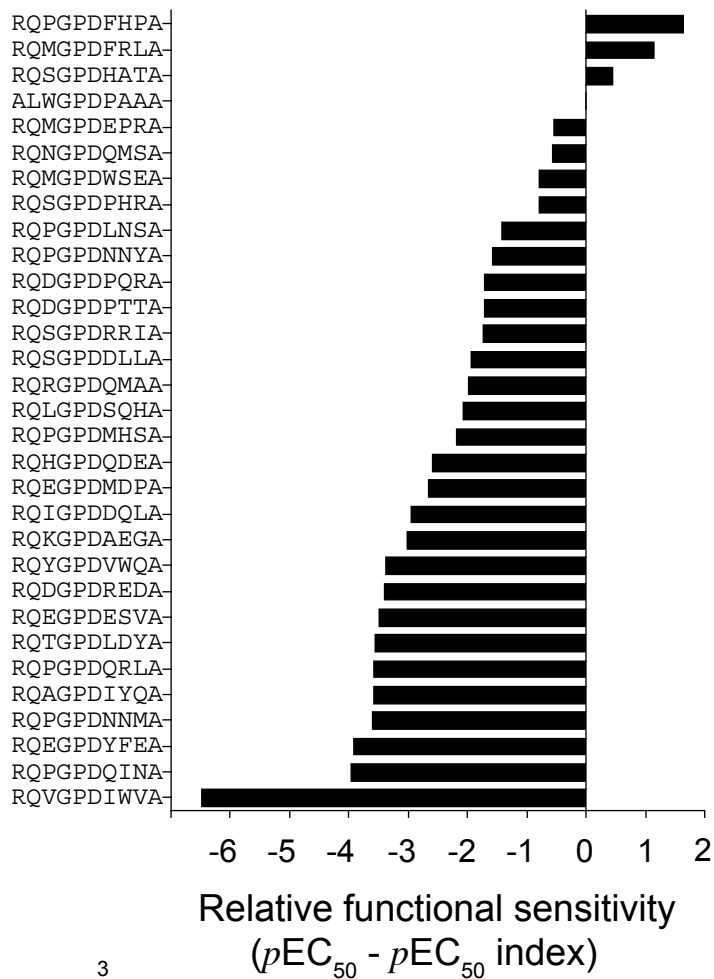
1. van den Berg, H. A., Rand, D. A., and Burroughs, N. J. (2001) *J Theor Biol* **209**, 465-486
2. van den Berg, H. A., and Rand, D. A. (2007) *Immunol Rev* **216**, 81-92

Figure S1

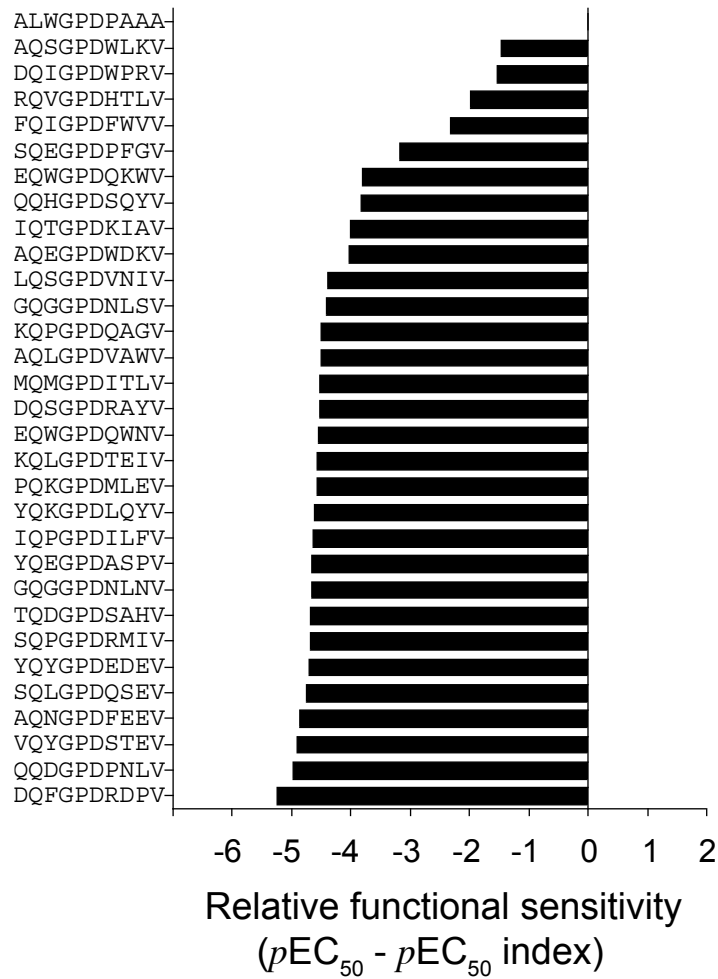




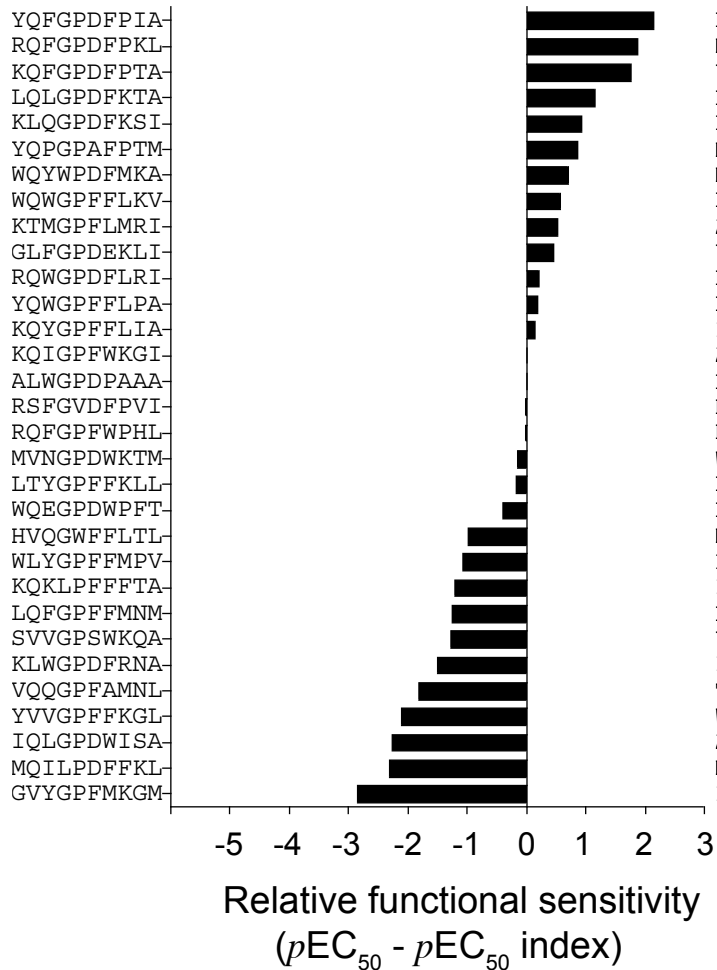
A



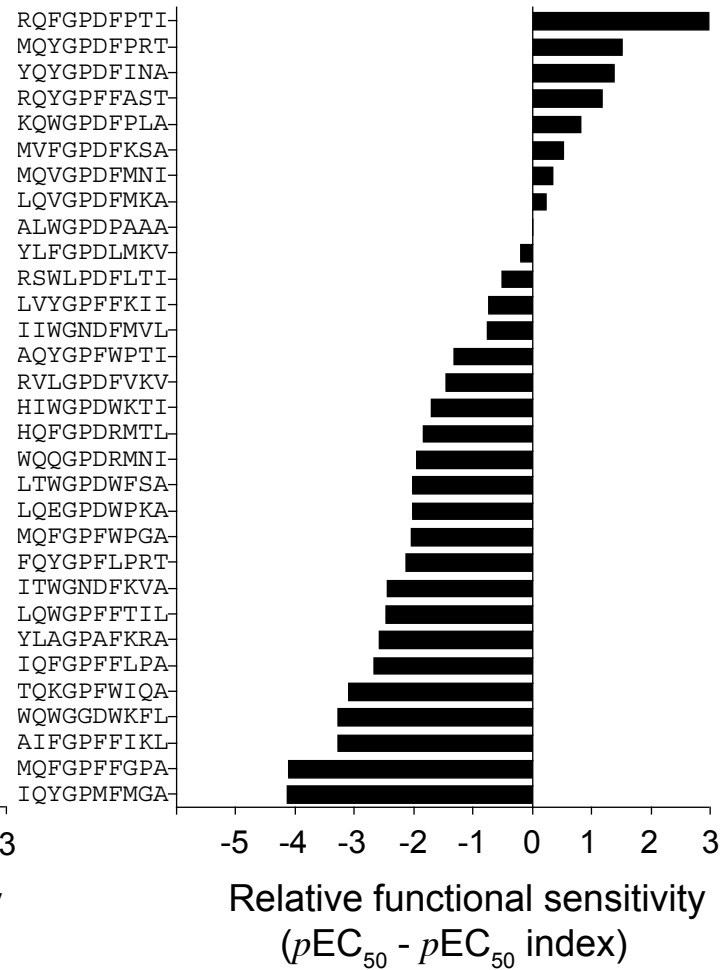
B



A



B



Supplementary Material: *Mathematica* Code

```

AmNum[AmAc_] :=
Module[{},
  First[First[
    Position[{"A", "C", "D", "E", "F", "G", "H", "I", "K", "L", "M", "N", "P",
      "Q", "R", "S", "T", "V", "W", "Y"}, AmAc]]]];

GenerateSample[mat_, merN_, samN_] :=
Module[{res, matnorm, biases, cnt, cntstop}, cntstop = 500 samN; cnt = 0;
  res = {1};
While[Length[Union[res]] < samN && cnt < cntstop, Print["Attempt"]; cnt++; res
= Transpose[Table[RandomChoice[mat[[p]] -> {"A", "C", "D", "E", "F", "G", "H",
"I", "K", "L", "M", "N", "P", "Q", "R", "S", "T", "V", "W", "Y"}, samN], {p,
merN}]]];
  Print["There are ", Length[Union[res]], " distinct peptides in the sample."];
  matnorm = Transpose[Table[Normalize[mat[[p]], Total], {p, merN}]];
  biases = Normalize[Exp[-Map[Total, Log[Table[matnorm[[AmNum[res][[i, p]]], p]],
{i, samN}, {p, merN}]]], Total];
Transpose[Append[Transpose[res], N[biases]]] // MatrixForm

EffectiveSampleSize[mat_, merN_] := Module[{matnorm, p, HH, Htot},
(* actually, effective population size *)
  matnorm = Table[Normalize[mat[[p]], Total], {p, merN}];
  HH = Table[0, {p, 1, merN}]; Htot = 0;
  For[p = 1, p < merN + 1, p++,
    HH[[p]] = Sum[EntroTerm[matnorm[[p, a]], {a, 1, 20}];
    Htot += HH[[p]];
  ];
  Exp[Htot]
]

EntroTerm[x_] := Module[{}, If[x > 0, x Log[1/x], 0]]

MakeWeightMatrix[motif_] :=
Module[{}], MM = Table[0, {i, 10}, {j, 20}]; motif[[1]];
  For[i = 1, i < 11, i++,
    MM[[i, First /@
StringPosition["ACDEFGHIKLMNPQRSTVWY", StringSplit[motif[[i]]]]] = 1]; MM]

Example of simultaneous fitting:

NumPep = 21; inits = Table[-6, {i, 3 + NumPep}]; inits[[1]] = 0;
inits[[2]] = 3000; inits[[3]] = 1; Print[inits]; pars = Array[a, 3 + NumPep];
PickPar[pars_, y_, ymax_] := Sum[pars[[3 + i]]*If[i == y, 1, 0], {i, ymax}];
model[pars_, SetInd_, NumPep_] := pars[[1]] + (pars[[2]] - pars[[1]])/(1 +
10^{(-pars[[3]])*(x - PickPar[pars, SetInd, NumPep])}); fitresult =
FindFit[data, model[pars, y, NumPep], Table[pars[[i]] (1 - j) + inits[[i]] j,
{i, 3 + NumPep}, {j, 0, 1}], {x, y}, MaxIterations -> 1000]

NumPerSet = 10; MyPlotRange = {{-14, -3}, {-150, 2000}}; plts =
Array[b, NumPep]; Do[
  plts[[i]] = Plot[model[pars, i, NumPep] /. fitresult, {x, -14, -1}, PlotStyle -
> Black, PlotRange -> MyPlotRange], {i, 1, NumPep}]; tmp_dat1 =
Show[plts[[Range[NumPep]]], PlotRange -> MyPlotRange, Frame -> True]; tmp_dat2
=ListPlot[data[[Range[NumPerSet NumPep], {1, 3}]], Frame -> True, PlotStyle ->
Black, PlotRange -> MyPlotRange]; Show[tmp_dat1, tmp_dat2]

```

Figure S6

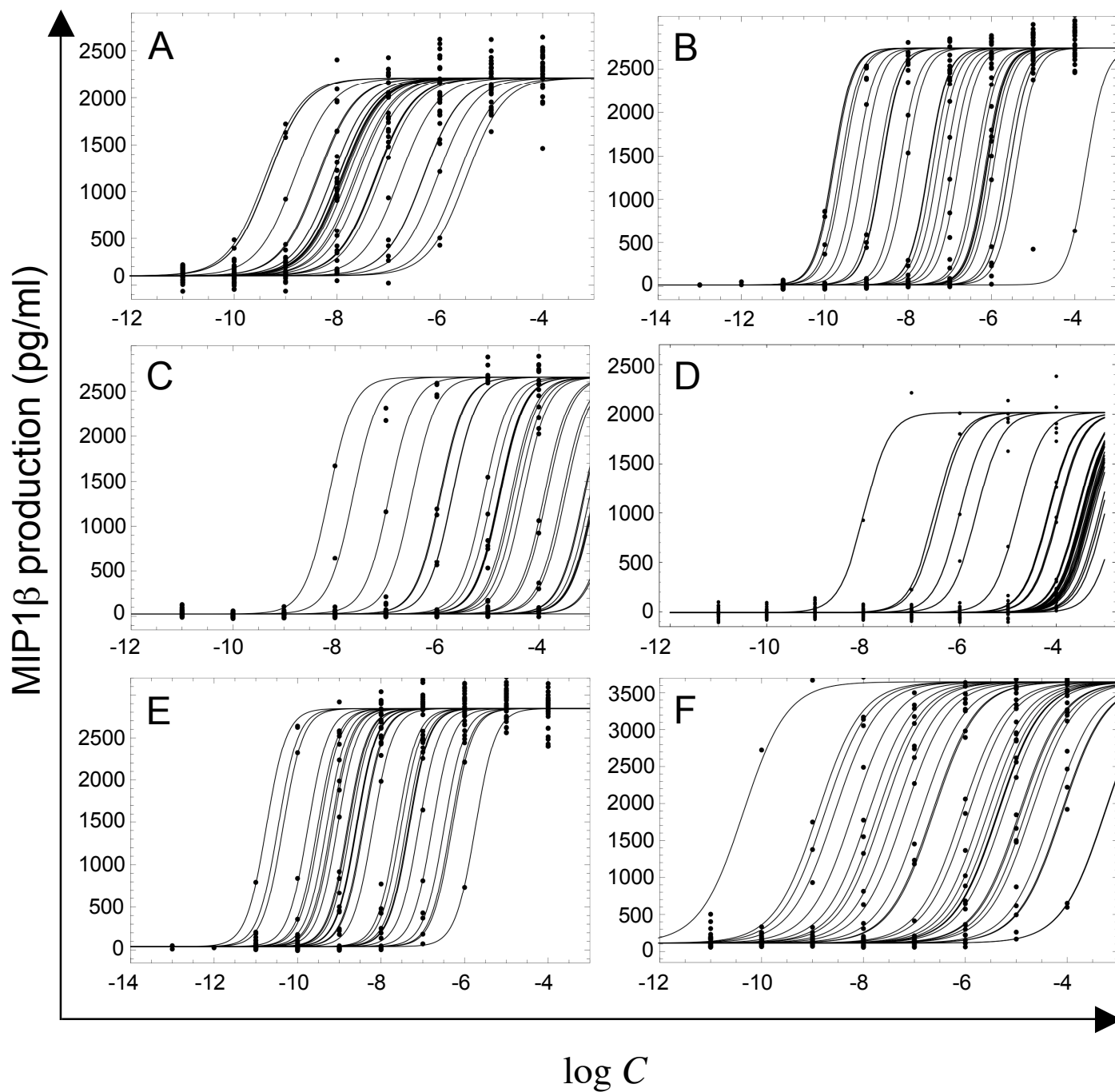


Table S1

Amino Acid

Peptide Position

	A	C	D	E	F	G	H	I	K	L
1	302.74	0	5	5	70.66	365.48	512.89	361.83	1528.32	891.17
2	5	0	5	5	5	110.86	5	246.09	5	949.64
3	400	0	5	400	2856.85	5	110.25	441.62	400	607.31
4	5	0	5	5	5	2234.31	5	5	5	50
5	5	0	5	5	5	5	50	5	5	5
6	398.38	0	3064.57	5	2271.47	5	5	5	5	5
7	121.83	0	15.84	10	3044.47	10	224.16	10	10	254.01
8	500	0	5	61.52	200	46.90	307.61	553.71	2861.12	1811.57
9	518.38	0	5	5	200	913.71	364.87	1497.26	1252.39	360
10	2248.32	0	5	5	5	5	5	1944.97	5	2000

	M	N	P	Q	R	S	T	V	W	Y
1	821.12	5	5	5	1863.35	194.31	176.04	236.35	991.07	1061.12
2	61.52	66.40	5	4851.17	5	275.94	546.40	835.74	5	5
3	31.68	400	116.95	421.52	5	5	5	400	2348.22	2248.93
4	5	5	5	5	5	5	5	5	50	5
5	5	50	2711.88	5	5	5	5	50	50	5
6	50	5	5	5	50	5	5	5	5	5
7	103.55	10	10	10	140.71	10	91.37	10	1242.64	28.02
8	2301.93	65.18	3363.05	5	200	5	5	200	18.88	276.55
9	31.07	908.83	1254.82	383.15	557.36	1237.16	1270.66	1195.74	5	15.84
10	345.99	5	17.06	5	5	5	854.01	1000	5	5

Table S2

Amino Acid

Peptide Position

	A	C	D	E	F	G	H	I	K	L
1	0.095 103	0	0.001 571	0.001 571	0.022 197	0.114 812	0.161 12	0.113 664	0.480 108	0.279 951
2	0.000 988	0	0.000 988	0.000 988	0.000 988	0.021 915	0.000 988	0.048 647	0.000 988	0.187 726
3	0.088 775	0	0.001 11	0.088 775	0.634 043	0.001 11	0.024 469	0.098 013	0.088 775	0.134 785
4	0.002 237	0	0.002 237	0.002 237	0.002 237	0.999 46	0.002 237	0.002 237	0.002 237	0.022 366
5	0.001 842	0	0.001 842	0.001 842	0.001 842	0.001 842	0.018 424	0.001 842	0.001 842	0.001 842
6	0.103 851	0	0.798 888	0.001 303	0.592 14	0.001 303	0.001 303	0.001 303	0.001 303	0.001 303
7	0.036 76	0	0.004 779	0.003 017	0.918 637	0.003 017	0.067 639	0.003 017	0.003 017	0.076 645
8	0.092 952	0	0.000 93	0.011 437	0.037 181	0.008 72	0.057 187	0.102 936	0.531 891	0.336 778
9	0.145 742	0	0.001 406	0.001 406	0.056 23	0.256 89	0.102 585	0.420 958	0.352 111	0.101 215
10	0.586 658	0	0.001 305	0.001 305	0.001 305	0.001 305	0.001 305	0.507 504	0.001 305	0.521 862

	M	N	P	Q	R	S	T	V	W	Y
1	0.257 945	0.001 571	0.001 571	0.001 571	0.585 352	0.061 042	0.055 301	0.074 245	0.311 333	0.333 339
2	0.012 162	0.013 125	0.000 988	0.958 982	0.000 988	0.054 548	0.108 012	0.165 209	0.000 988	0.000 988
3	0.007 03	0.088 775	0.025 957	0.093 552	0.001 11	0.001 11	0.001 11	0.088 775	0.521 159	0.499 123
4	0.002 237	0.002 237	0.002 237	0.002 237	0.002 237	0.002 237	0.002 237	0.002 237	0.022 366	0.002 237
5	0.001 842	0.018 424	0.999 297	0.001 842	0.001 842	0.001 842	0.001 842	0.018 424	0.018 424	0.001 842
6	0.013 034	0.001 303	0.001 303	0.001 303	0.013 034	0.001 303	0.001 303	0.001 303	0.001 303	0.001 303
7	0.031 246	0.003 017	0.003 017	0.003 017	0.042 458	0.003 017	0.027 57	0.003 017	0.374 954	0.008 455
8	0.427 936	0.012 117	0.625 202	0.000 93	0.037 181	0.000 93	0.000 93	0.037 181	0.003 51	0.051 411
9	0.008 734	0.255 52	0.352 796	0.107 723	0.156 703	0.347 829	0.357 249	0.336 184	0.001 406	0.004 453
10	0.090 279	0.001 305	0.004 45	0.001 305	0.001 305	0.001 305	0.222 838	0.260 931	0.001 305	0.001 305

Table S3

I: RQWGPDP{A/C/D/F/H/I/K/L/M/N/P/R/S/V/Y}
{A/C/G/H/I/K/L/M/N/P/Q/R/S/T/V}A motif

Number	Peptide Sequence	pEC ₅₀
1	RQWGPDPNCA	5.484
2	RQWGPDPCCA	5.651
3	RQWGPDPDEA	6.047
4	RQWGPDPDMA	6.319
5	RQWGPDPDCA	6.326
6	RQWGPDPDPA	6.777
Index	ALWGPDPAAA	7.151
7	RQWGPDPDRA	7.236
8	RQWGPDPDHA	7.24
9	RQWGPDPDKA	7.257
10	RQWGPDPDPA	7.473
11	RQWGPDPDKA	7.612
12	RQWGPDPDCA	7.659
13	RQWGPDPDFA	7.791
14	RQWGPDPDLA	7.854
15	RQWGPDPDPA	7.882
16	RQWGPDPDPA	7.884
17	RQWGPDPDKA	7.887
18	RQWGPDPDPA	7.907
19	RQWGPDPDPA	7.912
20	RQWGPDPDKA	7.916
21	RQWGPDPDPA	7.969
22	RQWGPDPDPA	7.979
23	RQWGPDPDPA	8.093
24	RQWGPDPDPA	8.096
25	RQWGPDPDPA	8.36
26	RQWGPDPDPA	8.384
27	RQWGPDPDPA	8.823
28	RQWGPDPDPA	9.328
29	RQWGPDPDPA	9.335
30	RQWGPDPDPA	9.395

II: RQWGP{D/F}{P/F}xx{A/I/L/V} motif

Number	Peptide Sequence	pEC ₅₀
1	RQWGPFFPSYV	3.723
2	RQWGPFFMDA	5.401
3	RQWGPFFIRI	5.561
4	RQWGPFFPDL	5.564
5	RQWGPFFRGA	5.658
6	RQWGPDFGEI	5.949
7	RQWGPFFFPWV	6.048
8	RQWGPFFRTA	6.102
9	RQWGPFFPHFV	6.121
10	RQWGPFFSQL	6.147
11	RQWGPDFEGV	6.306
12	RQWGPDPDDI	6.391
13	RQWGPFFKNV	6.795
14	RQWGPDFWGI	6.949
15	RQWGPFFKSA	7.114
16	RQWGPDFFYV	7.28
17	RQWGPFFPAYV	7.404
18	RQWGPFFHTA	7.498
19	RQWGPDPGTI	7.51
20	RQWGPDFSFV	8.05
21	RQWGPDFVNI	8.206
22	RQWGPDFAHV	8.625
Index	ALWGPDPAAA	8.639
23	RQWGPDPESI	8.716
24	RQWGPDPHEI	9.086
25	RQWGPDFFFV	9.253
26	RQWGPDPMAA	9.556
27	RQWGPDPDPA	9.622
28	RQWGPDPDPA	9.774
29	RQWGPDPDPA	9.802
30	RQWGPDFMTV	9.808

III: RQxGPDxxxA motif

Number	Peptide Sequence	pEC ₅₀
1	RQVGPDIWVA	Null
2	RQPGPDQINA	2.538
3	RQEGPDYFEA	2.567
4	RQPGPDNMA	2.891
5	RQAGPDIYQA	2.901
6	RQPGPDQRLA	2.92
7	RQTGPDLDYA	2.926
8	RQEGPDESVA	2.996
9	RQDGPDREDA	3.099
10	RQYGPDVWQA	3.12
11	RQKGPDAEGA	3.479
12	RQIGPDDQLA	3.536
13	RQEGPDMMPA	3.84
14	RQHGPDQDEA	3.894
15	RQPGPDMHSA	4.302
16	RQLGPDSQHA	4.409
17	RQSGPDQMAA	4.502
18	RQSGPDDLAA	4.548
19	RQSGPDRRIA	4.756
20	RQDGPDPPTA	4.771
21	RQDGPDPQRA	4.784
22	RQPGPDNNYA	4.922
23	RQPGPDLNSA	5.076
24	RQSGPDPHRA	5.694
25	RQMGPDWSEA	5.701
26	RQNGPDQMSA	5.924
27	RQMGPDPEPA	5.944
Index	ALWGPDPAAA	6.49
28	RQSGPDHATA	6.932
29	RQMGPDFRLA	7.642
30	RQPGPDFHPA	8.12

Table S3 cont

IV: xQxGPDxxxV motif

Number	Peptide Sequence	pEC ₅₀
1	DQFGPDRDPV	2.731
2	QQDGPDPNLV	2.991
3	VQYGPDSTEV	3.064
4	AQNGPDFEEV	3.109
5	SQLGPDQSEV	3.226
6	YQYGPDEDEV	3.261
7	SQPGPDRMIV	3.29
8	TQDGPDSAHV	3.294
9	GQGGPDNLNV	3.308
10	YQEGPDASPV	3.32
11	IQPGPDILFV	3.342
12	YQKGPDLQYV	3.37
13	PQKGPDMLEV	3.413
14	KQLGPDTEIV	3.417
15	EQWGPDQNVV	3.437
16	DQSGPDRAYV	3.45
17	MQMGPDITLV	3.457
18	AQLGPDVAWV	3.464
19	KQPGPDQAGV	3.477
20	GQGGPDNLSV	3.556
21	LQSGPDVNIV	3.577
22	AQEGPDWDKV	3.949
23	IQTGPDKIAV	3.975
24	QQHGPDSQYV	4.147
25	EQWGPDQKWV	4.166
26	SQEGPDFFGV	4.792
27	FQIGPDFWVV	5.654
28	RQVGPDHTLV	5.999
29	DQIGPDWPRV	6.446
30	AQSGPDWLKV	6.512
Index	ALWGPDPAAA	7.976

Va: Biased sampling set (1st)

Number	Peptide Sequence	pEC ₅₀
1	GVYGPFMKGM	5.741
2	MQILPDDFFKL	6.293
3	IQLGPDWISA	6.337
4	YVVGPFVKGL	6.487
5	VQQGPFAMNL	6.771
6	KLWGPDFRNA	7.095
7	SVVGPSWKQA	7.323
8	LQFGPFFMNM	7.343
9	KQKLPFFFTA	7.395
10	WLYGPFMPV	7.523
11	HVQGWFFLTL	7.617
12	WQEGPDWPFT	8.197
13	LTYGPFKLL	8.416
14	MVNGPDWKT	8.442
15	RQFGPFWPHL	8.582
16	RSFGVDFPVI	8.586
Index	ALWGPDPAAA	8.593
17	KQIGPFWKGI	8.601
18	KQYGPFFLIA	8.726
19	YQWGPFFLPA	8.78
20	RQWGPDFLRI	8.807
21	GLFGPDEKLI	9.043
22	KTMGPFLMRI	9.12
23	WQWGPFFLKV	9.158
24	WQYWPDFMKA	9.305
25	YQPGPAFPTM	9.461
26	KLQGPDFKSI	9.521
27	LQLGPDFKTA	9.749
28	KQFGPDFPTA	10.346
29	RQFGPDFPKL	10.467
30	YQFGPDFPIA	10.749

Vb: Biased sampling set (2nd)

Number	Peptide Sequence	pEC ₅₀
1	IQYGPFMFGA	3.254
2	MQFGPFFGPA	3.265
3	AIFGPFFIKL	4.093
4	WQWGGDWKFL	4.114
5	TQKGPFWIQA	4.279
6	IQFGPFFLPA	4.703
7	YLAGPAFKRA	4.81
8	LQWGPFFTIL	4.91
9	ITWGNDFKVA	4.939
10	FQYGPFLPRT	5.242
11	MQFGPFWPGA	5.346
12	LQEGPDWPKA	5.353
13	LTWGPDFWFA	5.365
14	WQQGPDRMNI	5.44
15	HQFGPDRMTL	5.553
16	HIWGPDWKTI	5.679
17	RVLGPDFVKV	5.931
18	AQYGPFWPTI	6.063
19	IIWGNDFMVL	6.618
20	LVYGPFFKII	6.646
21	RSWLPDFLTI	6.86
22	YLFGPDLMKV	7.186
Index	ALWGPDPAAA	7.381
23	LQVGPDFMKA	7.615
24	MQVGPDFMNI	7.716
25	MVFGPDFKSA	7.895
26	KQWGPDFPLA	8.203
27	RQYGPFFAST	8.564
28	YQYGPDFINA	8.754
29	MQYGPDFPRT	8.892
30	RQFGPDFPTI	10.357

A Single Autoimmune T Cell Receptor Recognizes More Than a Million Different Peptides

Linda Wooldridge, Julia Ekeruche-Makinde, Hugo A. van den Berg, Anna Skowera, John J. Miles, Mai Ping Tan, Garry Dolton, Mathew Clement, Sian Llewellyn-Lacey, David A. Price, Mark Peakman and Andrew K. Sewell

J. Biol. Chem. 2012, 287:1168-1177.

doi: 10.1074/jbc.M111.289488 originally published online November 18, 2011

Access the most updated version of this article at doi: [10.1074/jbc.M111.289488](https://doi.org/10.1074/jbc.M111.289488)

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2011/11/18/M111.289488.DC1.html>

This article cites 32 references, 11 of which can be accessed free at <http://www.jbc.org/content/287/2/1168.full.html#ref-list-1>