



GPU-Accelerated Discovery of Pathogen-Derived Molecular Mimics of a T-Cell Insulin Epitope

Thomas Whalley^{1,2†}, Garry Dolton^{1†}, Paul E. Brown³, Aaron Wall¹, Linda Wooldridge⁴, Hugo van den Berg⁵, Anna Fuller¹, Jade R. Hopkins¹, Michael D. Crowther^{1‡}, Meriem Attaf¹, Robin R. Knight⁶, David K. Cole^{1‡}, Mark Peakman⁶, Andrew K. Sewell^{1,2*} and Barbara Szomolay^{1,2}

¹ Cardiff University School of Medicine, Cardiff, United Kingdom, ² Systems Immunity Research Institute, Cardiff University, Cardiff, United Kingdom, ³ Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick Coventry, Coventry, United Kingdom, ⁴ Faculty of Health Sciences, University of Bristol, Bristol, United Kingdom, ⁵ Mathematics Institute, University of Warwick, Coventry, United Kingdom, ⁶ Peter Gorer Department of Immunobiology, Guy's Hospital, London, United Kingdom

OPEN ACCESS

Edited by:

Michael Loran Dustin,
University of Oxford, United Kingdom

Reviewed by:

Kathleen P. Pratt,
Uniformed Services University of the
Health Sciences, United States
Brian M. Baker,
University of Notre Dame,
United States

*Correspondence:

Andrew K. Sewell
sewellak@cardiff.ac.uk

[†]These authors have contributed
equally to this work

‡ Present address:

Michael D. Crowther,
Center for Cancer Immune Therapy,
Herlev Hospital, Copenhagen
University, Copenhagen, Denmark
David K. Cole,
Immunocore Ltd., Abingdon,
United Kingdom

Specialty section:

This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 27 September 2019

Accepted: 05 February 2020

Published: 28 February 2020

Citation:

Whalley T, Dolton G, Brown PE, Wall A, Wooldridge L, van den Berg H, Fuller A, Hopkins JR, Crowther MD, Attaf M, Knight RR, Cole DK, Peakman M, Sewell AK and Szomolay B (2020) GPU-Accelerated Discovery of Pathogen-Derived Molecular Mimics of a T-Cell Insulin Epitope. *Front. Immunol.* 11:296. doi: 10.3389/fimmu.2020.00296

The strong links between (Human Leukocyte Antigen) HLA, infection and autoimmunity combine to implicate T-cells as primary triggers of autoimmune disease (AD). T-cell crossreactivity between microbially-derived peptides and self-peptides has been shown to break tolerance and trigger AD in experimental animal models. Detailed examination of the potential for T-cell crossreactivity to trigger human AD will require means of predicting which peptides might be recognised by autoimmune T-cell receptors (TCRs). Recent developments in high throughput sequencing and bioinformatics mean that it is now possible to link individual TCRs to specific pathologies for the first time. Deconvolution of TCR function requires knowledge of TCR specificity. Positional Scanning Combinatorial Peptide Libraries (PS-CPLs) can be used to predict HLA-restriction and define antigenic peptides derived from self and pathogen proteins. *In silico* search of the known terrestrial proteome with a prediction algorithm that ranks potential antigens in order of recognition likelihood requires complex, large-scale computations over several days that are infeasible on a personal computer. We decreased the time required for peptide searching to under 30 min using multiple blocks on graphics processing units (GPUs). This time-efficient, cost-effective hardware accelerator was used to screen bacterial and fungal human pathogens for peptide sequences predicted to activate a T-cell clone, InsB4, that was isolated from a patient with type 1 diabetes and recognised the insulin B-derived epitope HLVEALYLV in the context of disease-risk allele HLA A*0201. InsB4 was shown to kill HLA A*0201⁺ human insulin producing β -cells demonstrating that T-cells with this specificity might contribute to disease. The GPU-accelerated algorithm and multispecies pathogen proteomic databases were validated to discover pathogen-derived peptide sequences that acted as super-agonists for the InsB4 T-cell clone. Peptide-MHC tetramer binding and surface plasmon resonance were used to confirm that the InsB4 TCR bound to the highest-ranked peptide agonists derived from infectious bacteria and fungi. Adoption of GPU-accelerated prediction of T-cell agonists has the capacity to revolutionise our understanding of AD by identifying potential targets for autoimmune T-cells. This approach has further potential for dissecting T-cell responses to infectious disease and cancer.

Keywords: type 1 diabetes, T-cell receptor, peptide-HLA, insulin, molecular mimicry, general-purpose computing on graphics processing units (GP-GPU), Compute Unified Device Architecture (CUDA), Nvidia

INTRODUCTION

T-cells protect against infections and neoplasms by scanning the host proteome for anomalies via peptides bound in major histocompatibility (MHC) molecules at the cell surface. The peptide specificity of T-cells is determined by the T-cell receptor (TCR), a highly variable heterodimeric protein generated by the somatic rearrangement of variable, joining and diversity gene segments and the quasi-random introduction/deletion of nucleotides to bestow additional non-germline-derived variation. Although it is estimated that this process can theoretically generate $\sim 10^{18}$ different $\alpha\beta$ TCRs in human (1) the actual repertoire possessed by any individual is believed to be of the order of 10^8 distinct TCR sequences (2, 3). The major variation between different TCRs is manifest in the antigen binding site, composed of six composite complementarity-determining region (CDR) loops, which contact both the peptide and the restricting MHC (4, 5). Effective immune coverage requires that the TCR repertoire must be capable of responding to almost all foreign peptide sequences that could be presented by host MHC molecules (1, 6). Unlike the B-cell receptor, the TCR is fixed and is not susceptible to further alteration in amino acid sequence or affinity maturation. There is no time to manufacture new TCRs in the face of an infection, so it is essential that the receptors on peripheral naïve T-cells are capable of dealing with any new challenge. Failure to cover all possible foreign peptide antigens would leave blind spots in T-cell immunity that pathogens could quickly evolve to exploit (1). Thus, successful immunity requires that a repertoire of $\sim 10^8$ TCRs be capable of responding to a vastly greater number ($>10^{16}$) of potential foreign peptides of a length and sequence capable of being presented by self MHC molecules (6). This evolutionary challenge is overcome by a phenomenon termed TCR degeneracy which refers to the ability of individual TCRs to transmit activation signals from very large numbers of different peptides bound in a single MHC molecule (7). The resulting T-cell crossreactivity, where a single T-cell clonotype can respond to a wide array of different peptide sequences, is thus an essential feature of T-cell immunity (1, 6).

Detailed examination of the potential for T-cell crossreactivity to trigger autoimmune disease (AD) will require means of predicting which peptides might be recognised by autoimmune TCRs. Here, we develop tools for this purpose that enabled the complex calculations involved in screening a wider pathoproteome to be undertaken by individual researchers on their own desktop computers. For this purpose, we focused on the AD type 1 diabetes (T1D) which arises through immune destruction of insulin-producing β -cells in the pancreas. T1D is triggered by the direct destruction of insulin-producing cells by cytotoxic T-cells. Four MHC class I molecules, human leukocyte antigen (HLA)*0201 (HLA A2 hereon), HLA A*24, HLA B*18, and HLA B*3906 are risk factors for T1D (8). Post mortem studies of patients close to disease onset have showed that numerous CD8 T-cells present in the characteristic islet mononuclear cell infiltrate

(9). Efficient adoptive transfer of disease in animal models typically requires CD8 T-cells and disease in such models is prevented by genetic modifications that abrogate MHC class I expression (10). Transgenic introduction of the HLA allele *HLA A2* into diabetes-prone non-obese diabetic (NOD) mice markedly accelerates disease development (11). More recent studies have described TCRs on activated CD8 T-cells in patient blood that can recognise β -cell-specific peptides presented through known disease risk HLA class I alleles (12–14) and such cells have been placed at the “scene of the crime” by *in situ* tetramer staining of insulinitic lesions (15). We previously searched for potential pathogen-derived peptide ligands for the 1E6 TCR which was isolated from a type 1 diabetic and shown to exhibit glucose-mediated killing of human HLA A2⁺ pancreatic beta cells (12, 14, 16). Two agonist peptides identified were mapped to the human pathogens *Bacteroides fragilis/thetaiotaomicron* and *Clostridium asparagiforme* demonstrating the potential utility of this approach for discovery of molecular mimicry (16, 17). These peptide agonists of sequence MVWGPDPPLYV and RQFGPDWIVA, respectively, differed from the wildtype insulin-derived sequence, ALWGPDPAAA presented at the cell surface of HLA A2⁺ human beta cells, and from each other, at seven of ten positions highlighting the potential extent of T-cell crossreactivity in this system. Pathogen-derived agonist peptides were identified using positional scanning combinatorial peptide library (PS-CPL) data to search the human pathoproteome (16).

Proteome database searches, in particular those derived from bacterial genomes, are computationally demanding on Central Processing Units (CPUs) and therefore of limited applicability. Here we sought to overcome this limitation using General-Purpose Graphics Processing Units (GP-GPUs) to reduce the execution time by a considerable degree. The great advantage of using a GP-GPU resides in the ability to parallelise the task at a massive scale, which allows for a huge increase in the number of sub-processes carried out per second (the throughput); even if the total time required to complete a sub-process (the latency) is somewhat increased, the net gains can be considerable. Compute Unified Device Architecture (CUDA), Nvidia's[®] GPU computing platform, is a free software that allows speed up of compute-intensive applications by harnessing the power of GPUs for the parallelisable part of the computation. CUDA has already been implemented in several bioinformatics tools ranging from molecular dynamics, sequence alignment, structure prediction, and spatio-temporal modelling (18–21). Many of these studies used consumer graphics cards aimed principally at gamers and costing no more than a few hundred pounds. CUDA provides a powerful and cost-effective way of accelerating computations that can be run on desktop setups without the need to access a high-performance computing (HPC) server (22). Use of CUDA and standard desktop computation means source code can be shared among researchers and run independently, allowing for modifications and greater ease of use; it also negates increases in wall time that would

be attributed to the nature of HPC, including job queueing and hosting.

We previously developed a peptide scoring algorithm that generates a ranked list of self and corresponding pathogenic viral peptides predicted to act as T-cell agonists from PS-CPL data generated with a given T-cell clone (17, 23). Here we extend this platform to include an expanded human pathoproteome by building protein databases from pathogenic bacteria and fungi and by implementing the peptide scoring algorithm into a GPU-accelerated framework which is suitable in view of the fact that our algorithm fits the single instruction on multiple data (SIMD) paradigm (24).

In summary, we present a GPU-accelerated strategy for epitope discovery and characterisation that uses the CPL-driven peptide scoring method developed by Szomolay et al. (17) to rank bacterial and fungal peptides occurring in very large curated databases in order of likelihood of recognition. Our GPU-implementation of the scoring algorithm significantly reduces the computational time and cost and has the potential to become an efficient application for ligand hunting. C++ is freely downloadable and can be run on any platform. GPU can be also included in most personal computers thereby democratising the approach as a publicly accessible executable, available at <https://github.com/WhalleyT/PICPL>. The executable contains both the C++ and CUDA-versions of the algorithm. It also contains a test data (a sample of protein sequences and CPL data) that the user can run on any of the three platforms (Windows, Linux, MacOS operating systems) and can be downloaded from the Releases.

MATERIALS AND METHODS

Cell Lines and T-Cells

HLA A2 restricted clone InsB4 that recognises the HLVEALYL peptide from the insulin chain (residues 10–18) (25) was grown from CD8 T-cells purified from a T1D patient using a T-cell library (26). Clone InsB6 was produced at the same time as InsB4 and subsequently found to express the same TCR. T-cell clones were routinely expanded using irradiated (3,000–3,100 cGy) allogenic PBMCs from three donors and 1 µg/mL of phytohaemagglutinin (PHA) (Alere, Thermo Scientific, Waltham, MA, USA) in T-cell media made from RPMI-1640 supplemented with 10% fetal bovine serum (FBS), 2 mM L glutamine, 100 U/mL penicillin, 100 µg/mL streptomycin, 10 mM HEPES 0.5X non-essential amino acids and 1 mM sodium pyruvate (all from Life Technologies, Carlsband, CA, USA) supplemented with 200 IU of IL-2 (Aldesleukin, Proleukin, Prometheus, San Diego, CA, USA) and 25 ng/mL human IL-15 (Peprotech, Rocky Hill, NJ, USA). T-cells were cultured for 2–4 weeks post-restimulation before being used for assays. K562 cell lines expressing HLA A2 and preproinsulin or GAD65 were generated and cultured as previously described (14, 27). CIRs expressing HLA A2 (28) were grown as suspension cells in R10 medium (RPMI-1640 with 10% FBS, 2 mM L-glutamine, 100 U/mL penicillin and 100 µg/mL streptomycin). Human pancreatic β-cells were produced as previously described (14, 27). T2 cells (HLA

A2⁺) were used as peptide presenting cells and cultured as for K562s.

T-Cell Functional Assays

Functional sensitivity assays using peptides were performed as previously described (29). T-cells were “rested” overnight in R5 (as for R10 but with 5% FBS) to reduce spontaneous release of the chemokine, MIP-1β. Individual peptides were synthesised to >40% purity [GL Biochem (Shanghai) Limited, Shanghai, China] and titrated for incubation overnight in R5 medium with the respective clone. Supernatants were harvested for MIP-1β ELISA using 96 well half area plates and a human MIP-1β DuoSet ELISA kit from R&D Systems (Minneapolis, MN, USA). Cytotoxicity assays using chromium 51 (17) or a non-radioactive europium hydrophilic ligand (14) were performed as described previously, with minor adjustments: for chromium release assays, 2,000 target cells were used per well. Data was initially analysed using GraphPad Prism software to generate the EC₅₀ of activation (concentration of peptide at 50% of maximum activation), which was used for further calculations as described below.

Sizing and Combinatorial Peptide Library (CPL) Scans

While MHC class I-restricted T-cells can recognise vast numbers of different peptides, all the strongest agonists tend to be of a single defined length of between 8 and 14 amino acids (30). The optimal length peptide recognised by InsB4 T-cells was determined by examining the response to six random peptide mixtures, X⁸, X⁹, X¹⁰, X¹¹, X¹², and X¹³ (where X is any of the 19 proteogenic L-amino acids excluding cysteine) in 96U well-plates in R5 medium (as for R10 but with 5% FBS) as for the functional peptide assays above and described elsewhere (30). MIP-1β was used as a readout as it is the most sensitive CD8 T-cell functional indicator (31, 32). Sizing and CPL scan peptide mixtures were purchased lyophilised at >40% purity (PepsScan Presto BV, Lelystad, The Netherlands) and stored as 80 or 20 mM DMSO stocks respectively, at –80°C. Working stocks (10 or 1 mM, respectively) were prepared in R0 (as for R10 but with no FBS), stored at 4°C, and used in assays at 1 or 0.1 mM, respectively.

T-Cell Receptor Sequencing and Production of Soluble TCR Protein

The TCRs of InsB4 and InsB6 were sequenced in-house using Sanger sequencing techniques as described previously (33). Individual TCR chains were expressed in *E. coli* as soluble inclusion bodies and refolded as described (34).

Peptide-MHC Multimer Staining

An optimal staining protocol was used for pMHC multimer staining as previously described (35), involving preincubation with 50 nM of the protein kinase inhibitor Dasatinib (36), followed by staining with 0.5 µg of PE conjugated pMHC tetramer in 50 µL of FACS buffer (2% FBS in PBS),

then incubation with an anti-PE unconjugated antibody (23). Subsequently, cells were stained for 5 min at RT with 2 μ L of a cell viability stain (Vivid; Life Technologies) that had been diluted 1:40 using PBS then without washing with anti-CD8-APC antibody (clone BW135/80, Miltenyi Biotech, Bergisch Gladbach, Germany). Data acquisition was performed on a BD FACS Canto II (BD Biosciences, Franklin Lakes, NJ, US) and analysed using FlowJo Software (TreeStar, Ashland, OR, US).

Manufacture of Soluble Proteins

Soluble TCR protein and biotinylated pMHC1 were manufactured as previously described (34). Briefly, codon-optimised InsB4 TCR α and β chains, HLA A2 heavy chain and β 2m chain were generated by Genewiz. All sequences were confirmed by automated DNA sequencing (Eurofins). TCR expression constructs were designed with a disulphide linked construct to produce the soluble domains (variable and constant) for both the α (residues 1–204) and β chains (residues 1–245) (34). The HLA A2 heavy chain (residues 1–248) (α 1, α 2, and α 3 domains), tagged, or not tagged with a biotinylation sequence, and β 2m (residues 1–100) were also cloned and used to make the pMHC1 complexes. The TCR α and β chains, the HLA A2 α chain and β 2m sequences were inserted into separate pGMT7 expression plasmids under the control of the T7 promoter (5). Competent Rosetta DE3 *E. coli* cells were used to produce the TCR α and β chains, HLA A2 heavy chain and β 2m in the form of inclusion bodies (IBs) using 0.5 mM IPTG to induce expression and protein were chemically refolded as described previously (37).

Surface Plasmon Resonance

SPR Equilibrium binding analysis was performed using a BIAcore T200TM equipped with a CM5 sensor chip as previously reported (38, 39). HLA A2-NLSALGIFST, derived from insulin-like growth factor 2 mRNA binding protein and recognised by a different TCR, was used as a negative control on flow cell 1. SPR kinetic analyses were carried out to determine the K_D values for the TCR, at 25°C. For all kinetic experiments, approximately 500 RUs of pMHC was coupled to the CM5 sensor chip surface. The TCR was then injected ten times at serial dilutions, from a concentration of 407 μ M, at 30 μ L/min. The K_D values were calculated assuming 1:1 Langmuir binding [$AB = B \cdot AB_{MAX} / (K_D + B)$] and the data were analysed using a global fit algorithm (GraphPad Prism).

GP-GPU Parallelisation of the Peptide Scoring Algorithm

The code developed by Szomolay et al. (17) was initially written in Matlab because its ease of use enables rapid prototype development and testing. As most of the time was spent scoring amino acid positions in a peptide against a lookup table, and this process could be performed as a matrix operation, we optimised the code and were able to obtain a 10-fold speedup vs. the original prototype algorithm. The code was parallelised via Matlab's Parallel Computing Toolbox and scaled up to

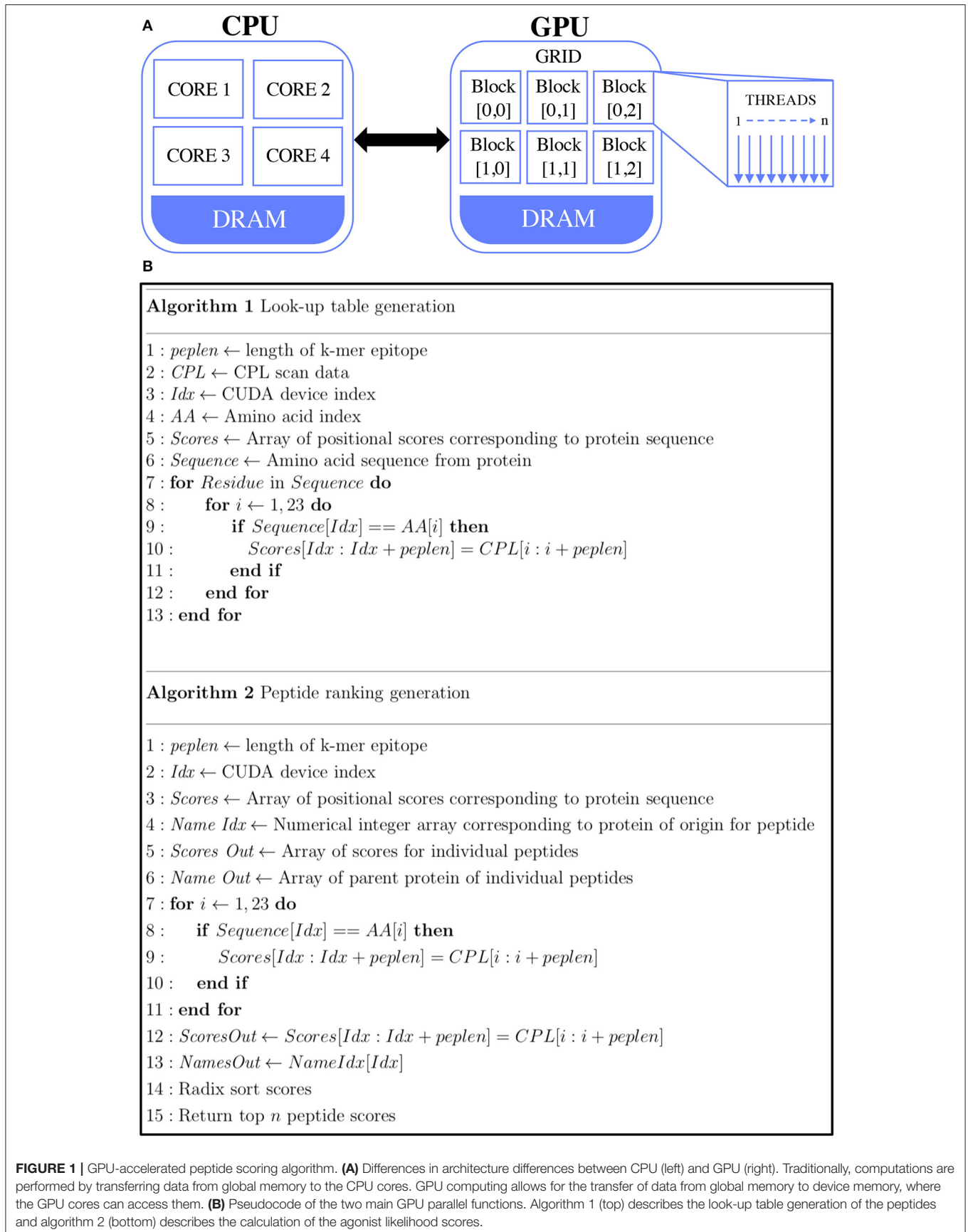
multiple hosts in a computational cluster using the Distributed Computing Server, which allows breaking the protein database up into sections and sending each section to a different thread. Results were then collated before the best overall peptide matches were found. The overhead of parallelisation using multithreading in Matlab is not significant and accordingly, the speed gain is virtually linear with respect to the number of parallel threads.

A more efficient approach to optimising the algorithm is to use GP-GPU programming, which takes advantage of GPU architecture that emphasises the parallel execution of very large numbers of relatively simple processes. The architecture differences between CPU and GPU are shown in **Figure 1A**. One operation is performed on many datasets (e.g., our protein databases are broken up into 30,000 protein sequences at a time and a single thread is launched for every protein sequence, thus analysing the data across multiple GPU blocks). Although Matlab provides a simple way to develop a prototype as the Parallel Computing Toolbox supports GP-GPU, substantial gains could be achieved by rewriting the program in C++. The code was optimised to minimise bottlenecks in serial processing and was adapted to run effectively on the GPU. This meant performing code refactoring to minimise memory allocation and transfer between the CPU and GPU by unrolling loops and passing every sequence together as a single one-dimensional array.

The look-up table and CPL algorithms are shown in **Figure 1B**. The CUDA code contains two main functions. The first function is the generation of the CPL scan data look-up table (Algorithm 1). In this, for every given amino acid a single thread was launched. This look-up table is kept in the GPU memory, where every k-mer is scored by its amino acid composition and positional score (Algorithm 2) which is done with one thread for every sequence. Here, the GPU RAM becomes a limiting factor. In the case of the Nvidia[®] Quadro K1200, this is 4GB of memory, which equates to about 30,000 sequences, along with the corresponding scoring data required. Once every potential agonist is scored in every given sequence, the scores are sorted in descending order based on their agonist likelihood score (Λ), a concept introduced in our previous study (17). The Λ -score makes use of the CPL scan data which essentially represents the recognition preference of a T-cell at each residue position, on average, over all its agonists. The peptide ranking is performed with the Thrust library of Bell and Hoberock (40). An in-built radix sort is used to sort the peptides, as this has been proven to be well-implemented in the CUDA code. An output file containing the score, peptide and parent sequence name are returned.

Benchmarking

Benchmarking was performed on four different desktop implementations of the algorithm: a single core implementation, a 4 core OpenMP parallel, an 8 core OpenMP implementation and a CUDA implementation. The CUDA libraries plus compiler and OpenMP are freely downloadable from <https://developer.nvidia.com/cuda-toolkit> and <https://www.openmp.org>, respectively.



Two benchmarking tests were performed to investigate the execution time. The first test was undertaken by running each implementation of the code across databases of varying sizes; ranging from 10^6 to 2×10^7 sequences. For consistency, the database was an artificially generated set of protein sequences 311 amino acids in length (the average protein length found in the bacterial database). The number of results returned and the CPL data used were the same in each test. The second test varied the number of results returned (i.e., the number of top-scoring peptides), ranging from 100 to 10^6 against a pathogenic fungal database consisting of 1.5×10^6 sequences. All code was benchmarked on a machine running Ubuntu GNOME 16.10, with a Nvidia® Quadro K1200 graphics card with 512 cores, an Intel Core 6700-k processor and 16 GB of memory. All GP-GPU applications were developed using CUDA, hence were compiled with the nvcc compiler (v8.0.44). The sequential C++ code was compiled using the GNU g++ compiler (v5.4.0) with the C++11 ANSI standard and using the -O1/-O2/-O3 parameter option. OpenMP parallel code was compiled with the same GNU compiler, with an OpenMP 4.0 compiler directive.

From Databases to Peptide Universe

The search space of the protein databases represents a tiny subsection of all possible peptides. For a given 9-mer, there are 20^9 (5.1×10^{11}) different unique amino acid combinations. This dwarfs the 16 billion peptides observed in the full bacterial universe. As well as the database-driven peptide-scoring algorithm, we have developed an optimised algorithm to score a selected peptide against the *entire* peptide universe for given a CPL data. The code once again leverages the CUDA library to optimise for speed. The *in silico* search utilises the GPU's architecture to parallelise in 3 dimensions. Therefore, every possible peptide is generated by means of collecting all sub-peptides a third of the total peptide size (e.g., for a 9-mer this would be $8,000 \times 3$ sets of 3-mers); which are then combined across all permutations on the GPU. The CUDA- and C++-version of the algorithm is available at <https://github.com/WhalleyT/PICPL>, within the downloadable executable.

RESULTS

T-Cell Clone InsB4 Recognises a *bona fide* HLA A2-Restricted Insulin Epitope Presented at the Surface of Human β -Cells

T-cell clone GD.InsB4 (InsB4 from hereon) was generated from the blood of a patient with T1D using the T-cell libraries approach and the putative human insulin β chain_{10–18} epitope, HLVEALYLV, as previously described (26), with peptide activation data shown in **Figure 2A**. This clone was initially shown to recognise K562 cells transduced with HLA A2 and preproinsulin (so-called surrogate human β -cells), which were not killed by a T-cell clone with a different specificity (**Figure 2B**). Killing of the surrogate cells by InsB4 was dependent on expression of preproinsulin as K562 cells transduced HLA A2 alone or with HLA A2 and GAD65 were not recognised (**Figure 2C**). These results show that the HLVEALYLV epitope

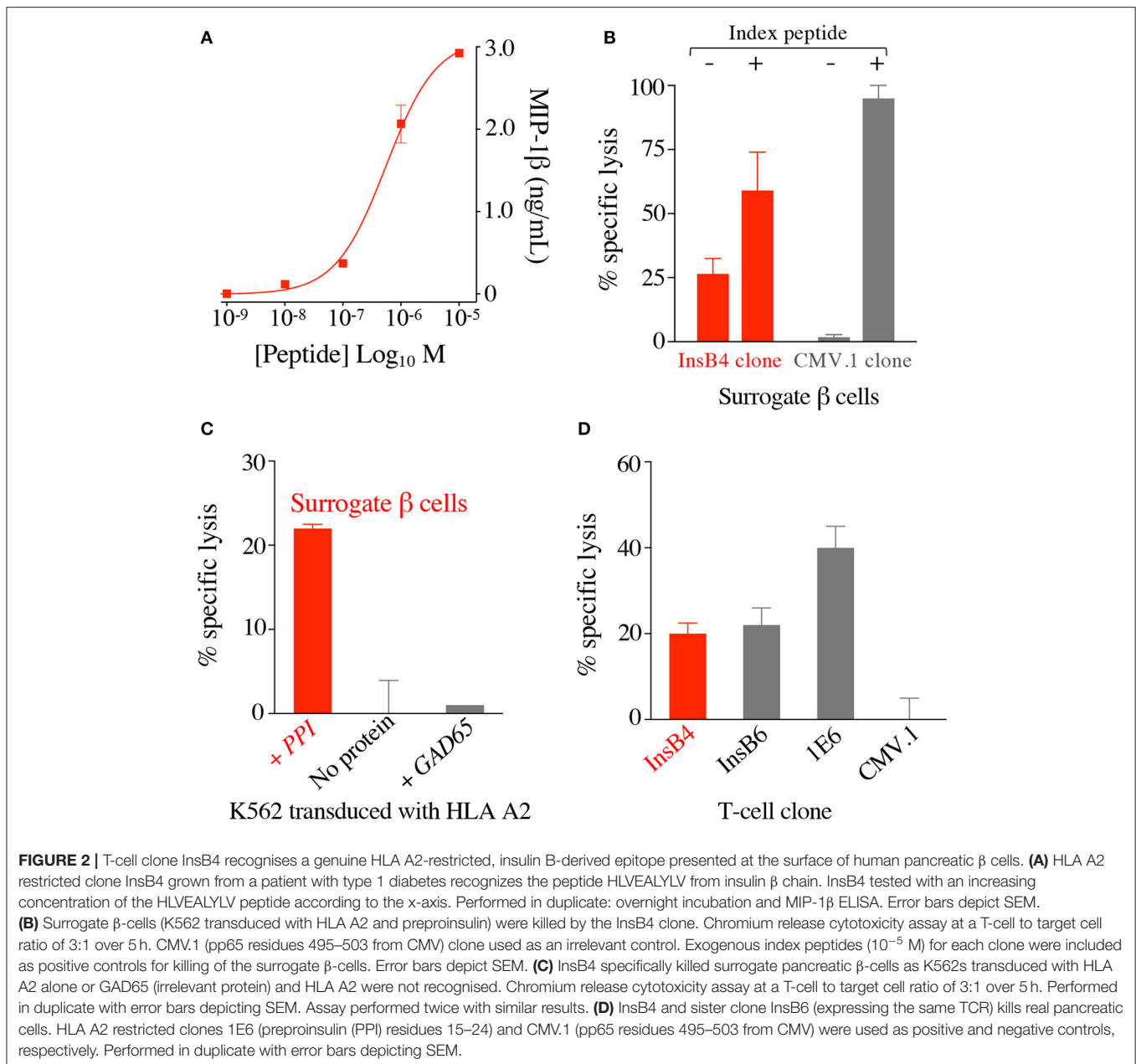
can be genuinely processed and presented at the cell surface in the context of HLA A2. We next demonstrated that InsB4, sister clone GD.InsB6 (InsB6) and preproinsulin-specific T-cell clone 1E6 (14) could lyse real human pancreatic β -cells (**Figure 2D**). A control clone, CMV.1 specific for the CMV pp65-derived epitope NLVPMVATV, did not kill these pancreatic cells in parallel assays. We conclude that InsB4 and InsB6 clones, derived from the blood of a patient with T1D can destroy human pancreatic β -cells and therefore have capacity to cause T1D. Importantly, these data represent the first time that the HLVEALYLV peptide has been confirmed as a genuine human β -cell epitope that is processed and presented by HLA A2⁺ pancreatic cells and capable of resulting in destruction of insulin producing cells.

CPL Screening of T-Cell Clone InsB4

Preference of InsB4 for nonamer peptides was confirmed by performing a sizing scan using random peptide mixtures of 8, 9, 10, 11, 12, and 13 amino acids in length (**Figure 3A**). This sizing scan indicated that the InsB4 T-cell exhibited a strong preference for peptides of 9 amino acids in length. We next undertook nonamer combinatorial library screening of the InsB4 clone. MIP1 β amounts produced by each of the 180 sublibraries in the PS-CPL scan is shown in **Figure 3B** and summarised as a logo plot in **Figure 3C**. The “index” glutamic acid sub-library in position 4 was recognised most strongly, with virtually no evidence that strong agonists display any other amino acid residue at this position. The strong preference for a glutamic acid residue at position 4 suggests that this residue may form the major TCR contact and, moreover, that the InsB4 TCR might focus on a precisely delimited hotspot. Sub-libraries with index peptide residues (i.e., those residues found in the HLVEALYLV insulin-derived peptide) at positions 1, 5 and 6 were recognised poorly. Selected sub-libraries at positions 1, 5, 6, 7, and 8 produced more MIP1 β than the index sub-library suggesting that it might be possible to create super-agonist peptides for this TCR that exhibit greater functional sensitivity than the index peptide sequence.

Enhanced Screening of Curated Pathogen Databases Using GPUs

We assembled databases containing protein sequences from fungi and bacteria. We have previously collated a database of viral proteomes of species that infect, or might infect, humans (17). The bacterial proteome database was subdivided into two pathogenicity classes: category I (species not known to cause, or which have been proven *not* to cause, disease in humans or animals) and category II (species known to infect humans or animals, including species that can cause human disease). For a given CPL scan, evaluation of 1,573,341 distinct, proteins from fungal pathogens and 11,838,978 distinct category I and 11,589,864 distinct category II bacterial proteins took 2, 12, and 13 min, respectively, on the GPU (outputting the 2000 top-scoring peptides in each case). The links to the publicly available fungal and bacterial databases are provided in the **Supplementary Material**.

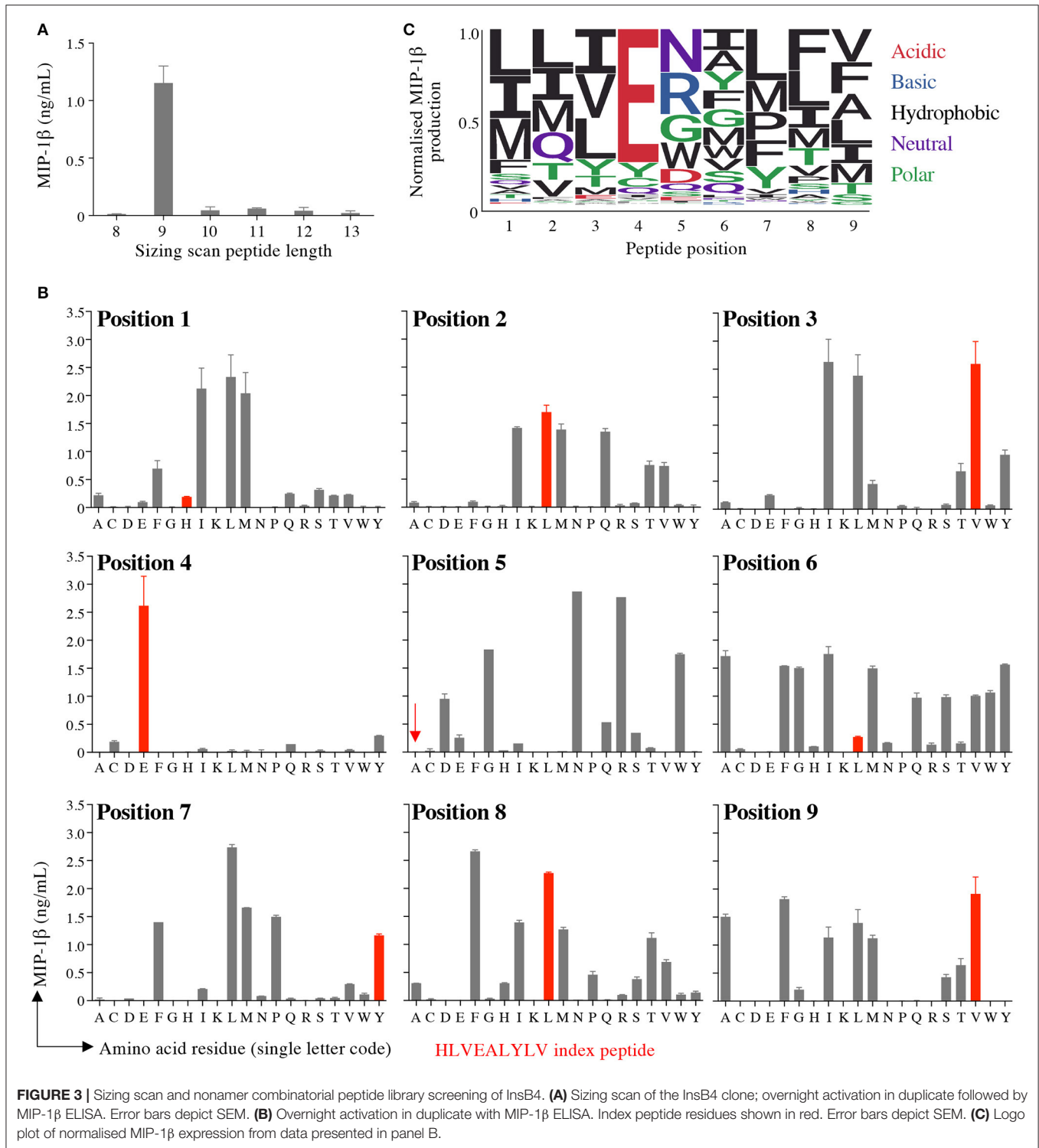


GPU Computing Efficiency

We evaluated the training performance of GPU acceleration on the number of top-ranked peptides and datasets of varying sizes and compared its performance with other platforms. When compared with the number of top-ranked peptides, the CUDA implementation of the algorithm was seemingly unaffected by the number of top-scoring peptides, unlike the other platforms (**Figure 4A**). GPU was superior when tested on datasets of different sizes and reduced the time to scan datasets by a factor of ~ 4.5 compared to C++ (**Figure 4B**) thus giving a much broader search depth. **Figure 4C** shows the average speed-up for datasets of different sizes. The CUDA implementation analysed an average of 18,589.4 sequences per second compared to 4,624.2

sequences and 11,800.6 sequences per second in the serial (C++) and OpenMP implementations, respectively. This results in a marked reduction of run time, which is conserved across varying input data sizes.

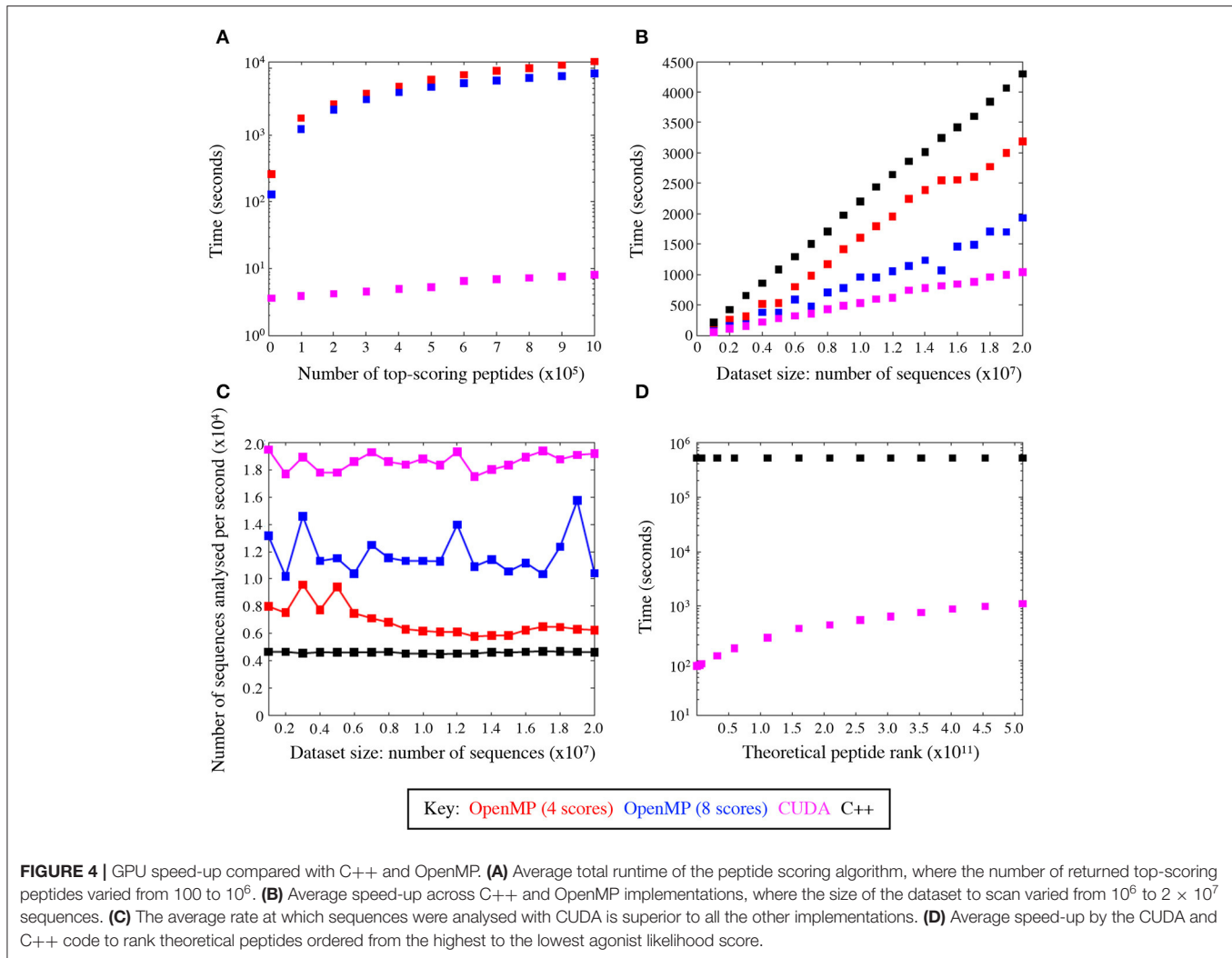
The CUDA code was also profiled using the Nvidia[®] profiler (<https://developer.nvidia.com/nvidia-visual-profiler>) which demonstrated that the GPU parallelisation removed the scoring function as the main bottleneck, leaving input/output operations as the limiting factor on the run time. This was evident from the Nvidia[®] profiler output, which attributed input/output operations (i.e., reading and writing the files to and from disk) to take up 28% of the code's run time. The functional CUDA code took 11% of the time with the rest being attributed to CPU



dependent operations. This shows that even with substantial further optimisation, the main time losses would be incurred with operations that are not easily relegated to the GPU.

Figure 4D compares the speed-up between the GPU and C++ codes scanned against the entire peptide universe. When comparing the two extreme cases for the highest (LLIENILFV)

and the lowest (GGCAADCDC) scoring possible peptides, we observed a significant increase in time. This is because every time the queried peptide scores lower than a given peptide the GPU threads must be synchronised and halted while memory is accessed to add to the ranking value. The highest scoring peptide was ranked in 82 s, whereas the lowest scoring peptide was ranked



in 1,117 s leading to >10-fold increase in runtime. Nevertheless, the GPU-accelerated algorithm performed significantly better than its serial C++ implementation, which took 16,088 and 16,144 s for the best and worst case, respectively. The index insulin-derived peptide (HLVEALYLV) ranked at 789,930,969 (still within the top 0.15% of all nonamers) in 85 s.

Validation of the GPU-Accelerated Scoring Algorithm With Randomly Chosen Uniformly Distributed Pathogen-Derived Peptides

We tested the GPU-accelerated strategy by interrogating the fungal and bacterial pathoproteome for peptides predicted to activate the autoimmune InsB4 clone. The PS-CPL data depicted in **Figure 3** were used to screen the protein databases compiled from fungal or bacterial species, as described above. As part of the database screening, a searching algorithm generates a hierarchal list of possible agonists by assigning each peptide with an agonist likelihood score (Λ) developed in (17). We previously demonstrated that the cognate peptide for antiviral

or antitumor clones could be identified using PS-CPL data, pathogenic viral (10,733 proteins) or human self (54,886 proteins) databases and the scoring algorithm using CPU, with a strong association between the Λ -value, and the peptide's functional sensitivity to activate the respective clone (17). Here, our objective was to establish if the GPU-accelerated algorithm could identify potential agonists from the pathogenic fungi database (1,573,341 proteins) and the entire pathogenic bacterial database (23,428,842 proteins).

For the purpose of validation, 72 fungal and 72 bacterial peptides were chosen, which were uniformly sampled without repetition and whose Λ -values spanned ~ 6 orders of magnitude, thus placing 12 peptides in six "bins," with each bin representing an order of magnitude. To achieve a sufficiently large span of unique Λ -values for sampling, the top 3×10^5 fungal and top 5×10^6 bacterial peptides were ranked by the searching algorithm based on the CPL data. This was only feasible by using multiple cores on GPUs, which increased the speed of peptide identification from >1 day to under 3 and 11 min, for the fungal and bacterial sequences, respectively.

Each peptide was tested for functional sensitivity in titration assays, with InsB4 activation quantified by MIP-1 β ELISA (Figures 5, 6). For the fungal protein-derived peptides it was evident that the most potent agonists resided in Bin 1 with 7 sequences being more potent than the index insulin_{10–18} sequence (Figure 5). Average ligand functional sensitivity decreased through Bins 2–5 such that activation was only observed with 4/12 peptides in the 5th Bin. Results were similar for peptides identified from the much larger bacterial protein database (Figures 6, 7A) except the greater number of overall peptides resulted in more being found that were of higher functional sensitivity than the index insulin_{10–18} sequence. The most potent sequences found were LLVENIPLF (Bin 1), IILEGILIL from (Bin 2), and ILLEDGMLL (Bin 3). None of the peptides in Bin 6 acted as agonists of the InsB4 T-cell (Figures 6, 7A).

Overall, it was apparent that peptides that scored as being most likely to be seen by InsB4 were indeed better agonists for InsB4, with fewer peptides being recognised when moving from Bin 1 to Bin 6, for both the fungal and bacterial peptides (Figure 7A). In order to make comprehensive comparisons between the 144 peptides, the pEC_{50} (defined as $-\log_{10} EC_{50}$) of InsB4 activation for every peptide was compared to that of the HLVEALYLV index (pEC_{50}^{index}), giving each peptide a relative functional sensitivity ($\Delta pEC_{50} = pEC_{50} - pEC_{50}^{index}$).

We examined the statistical association between the agonist likelihood score Λ and the relative functional sensitivity ΔpEC_{50} for all 144 randomly chosen peptides using both the Spearman's and Pearson's correlation coefficient (Figure 7B). For the fungal peptides, Spearman correlation was 0.42 (significant at the 0.5% level) and the Pearson correlation was 0.48 (significant at the 0.1% level). For the bacterial peptides, Spearman correlation was 0.36 (significant at the 0.1% level) and the Pearson correlation was 0.39 (significant at the 0.1% level). These results demonstrate that the CPL-driven database searching can accurately identify peptides derived from cellular organisms that were recognised by an autoimmune T-cell clone.

Pathogen-Derived Peptides Can Act as Super-Agonist Ligands for an Insulin B-Specific T-Cell Clone

Our CPL-based peptide searching algorithm identified several peptides derived from bacterial and fungal proteins that shared a very low sequence homology with the index insulin-derived peptide but were recognised by the InsB4 T-cell clone with higher functional sensitivity. Whereas, the above approach randomly sampled peptides from the bacterial and fungal databases, we next explored the top scoring peptides from each database and functionally tested 20 peptides. As shown in Figure 8, the top 5 most potent fungal InsB4 ligands were peptides derived from *Candida albicans* (MIVENVPLL), *Saccharomyces cerevisiae* (LIENAPLI), *Aspergillus niger* (LLVENWPLL), *Mucor circinelloides* (MLVEGVLLA), and *Sporothrix schenckii* (MIVEGFLLL). As shown in Figure 9, all 20 bacterial peptides

for InsB4 were recognised (19 gave unambiguous EC_{50} values), with 7 of the peptides eliciting greater activation from InsB4 than the index HLVEALYLV insulin-derived peptide. Furthermore, peptides from *Streptococcus caballii* (IIIEGILFV) and *Helicobacter pylori* (MLLENGLLA) acted as super agonists of InsB4. These data strongly suggest that the InsB4 TCR can bind to pathogen-derived peptides with higher affinity than with the index HLVEALYLV sequence that can mediate destruction of human pancreatic β -cells.

In order to formally prove enhanced TCR binding we examined pMHC tetramer staining and ligand interaction with soluble InsB4 TCR by SPR. The InsB4 T-cell clone stained well with tetramers made with HLA A2-MIVENVPLL (*C. albicans*) and HLA A2-MLLENGLLA (*H. pylori*) (Figure 10A). In contrast staining with the index peptide, insulin-derived HLA A2-HLVEALYLV tetramer was very weak. Peptide-HLA tetramer staining intensity is known to correlate with TCR affinity for the cognate ligand (29). Autoimmune TCRs are known to stain poorly with cognate tetramer (35, 41), so this weak staining was not unexpected. Interaction between the InsB4 TCR and HLA A2-HLVEALYLV was too weak to measure by SPR. In contrast, *C. albicans*-derived HLA A2-MIVENVPLL and *H. pylori*-derived HLA A2-MLLENGLLA antigens bound with a K_D of 60.7 and 168.5 μ M, respectively, confirming that these peptides engage the TCR within an affinity range capable of triggering T-cell activation (Figure 10B). Attempts to gather structural data of the InsB4 TCR in complex with a cognate antigen failed as we were unable to grow crystals of this TCR, despite multiple attempts.

DISCUSSION

There is a pressing need for efficient computational methodology for prediction of peptide mimics and super-agonists for therapeutically relevant T-cell clones. Here, we first validated a human insulin β chain_{10–18} epitope HLVEALYLV using T1D patient-derived T-cell clone InsB4 and showed that this T-cell could kill human pancreatic β -cells. These data represent the first time that the HLVEALYLV peptide has been confirmed as a naturally processed β -cell-presented epitope. Recognition patterns of random peptide mixtures of 8, 9, 10, 11, 12, and 13 amino acids in length clearly showed that the InsB4 T-cell preferred nonamer peptides. In order to investigate the peptide degeneracy profile of InsB4 we performed a nonamer PS-CPL screen. InsB4 demonstrated a clear preference for a glutamic acid residue at position 4 strongly suggesting that this residue acts as the primary TCR contact within the peptide. In support of this notion, this residue was prominent in all the super-agonist peptides we extracted from the pathoproteome.

In order to determine whether the InsB4 T-cell could cross-react with pathogen-derived peptides we turned to computational screening of PS-CPL data. *In silico* searching of the known terrestrial proteome with our algorithm previously developed using Matlab (17) takes several days on an HPC server. We reduced the time for these searches to under 30 min using GP-GPU allowing this approach to be run on a desktop or laptop computer. The search algorithm generates

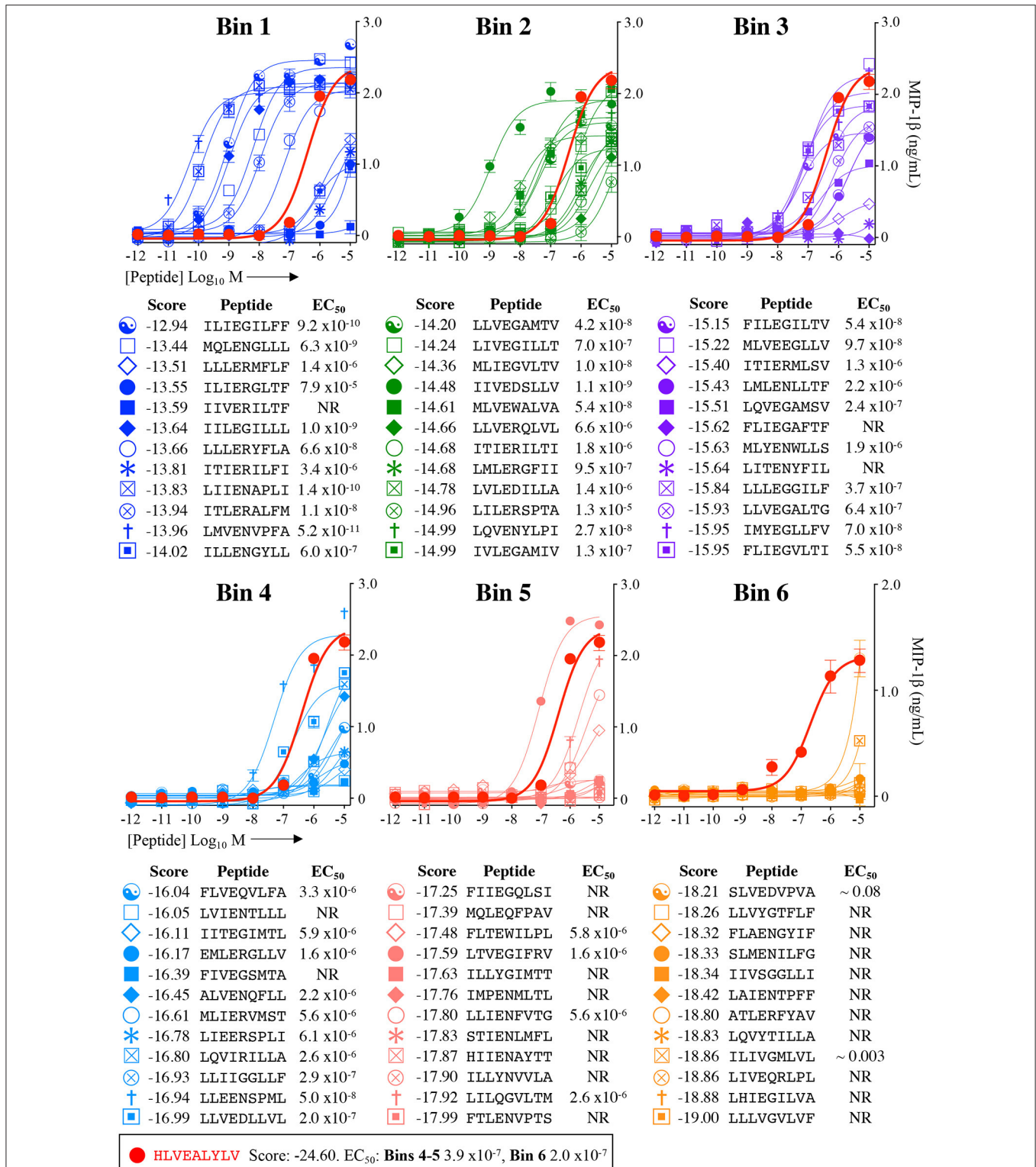


FIGURE 5 | Recognition of 72 randomly chosen and uniformly distributed fungal peptides by the InsB4 T-cell clone. Each panel shows results of MIP-1β assay, indicating activation of T-cell clone InsB4 by 12 peptides and the index HLVEALYLV peptide (shown as a bold red line). InsB4 was incubated overnight with the indicated peptides and MIP-1β released into the supernatants quantified by ELISA. Bins 1–5 were performed in the same assay with Bin 6 performed in a separate experiment hence the different EC₅₀ values for the index peptide. EC₅₀ for each peptide (M) is displayed in the tables. NR, no T-cell response. The mean pEC₅₀ for each bin is plotted in **Figure 7A**.

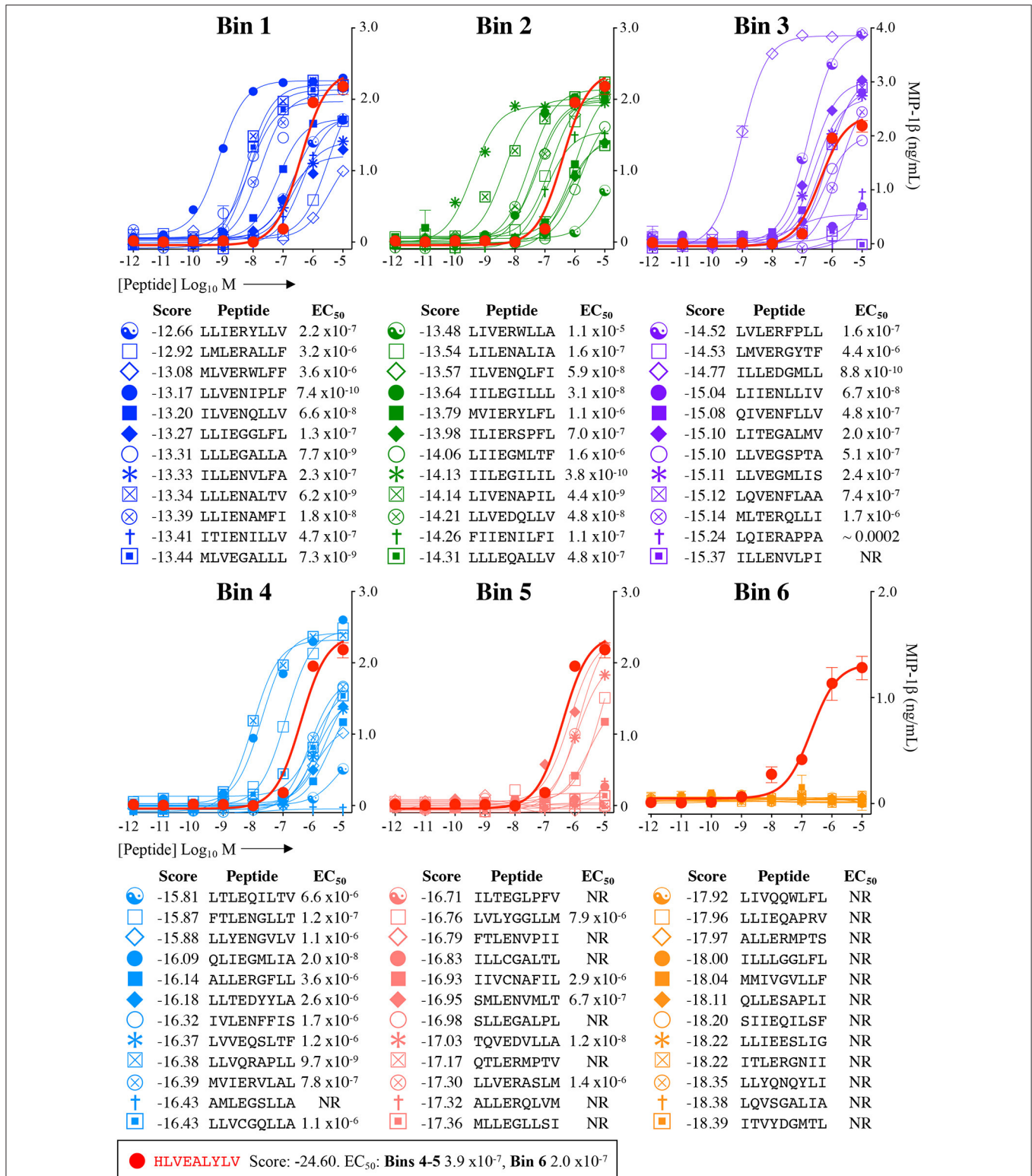
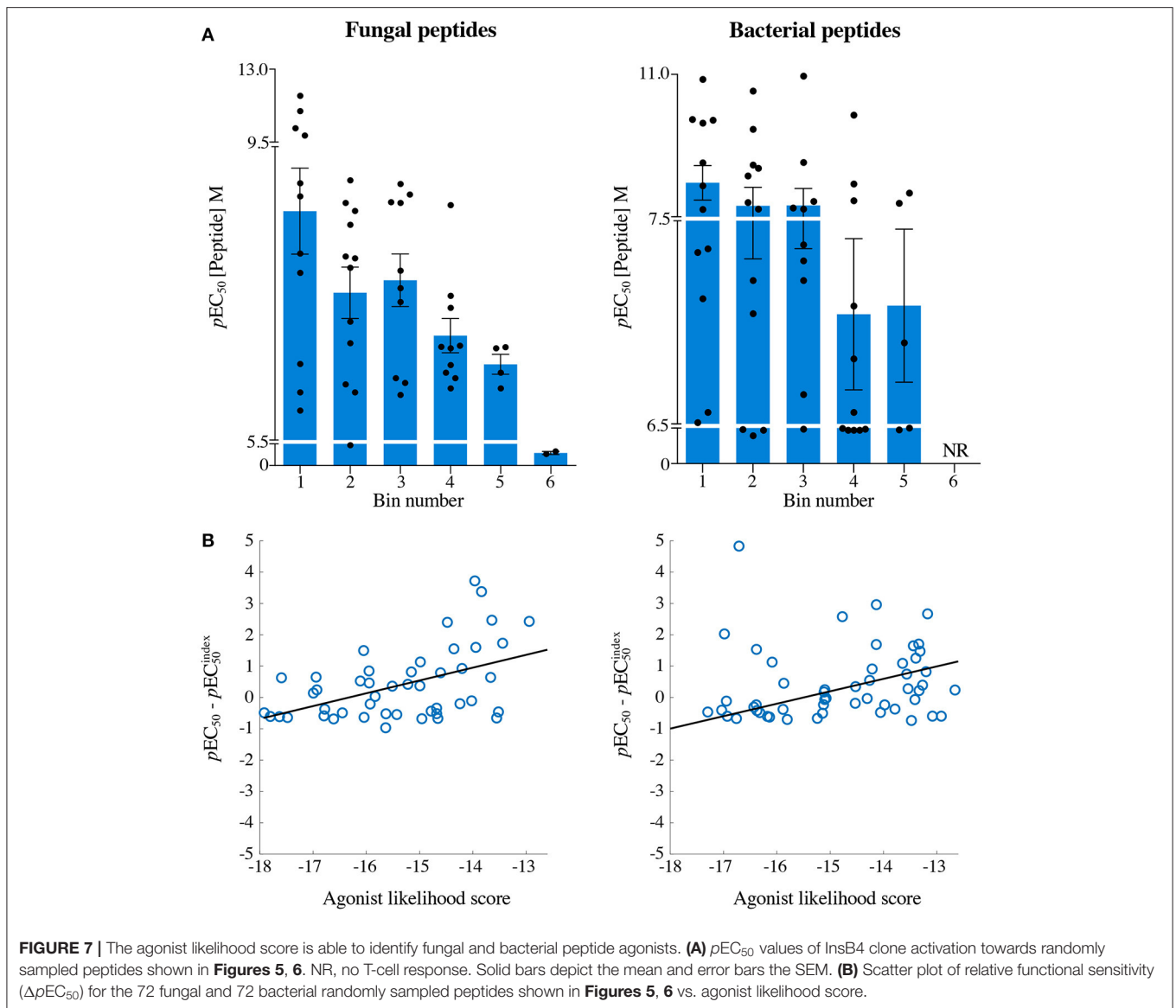


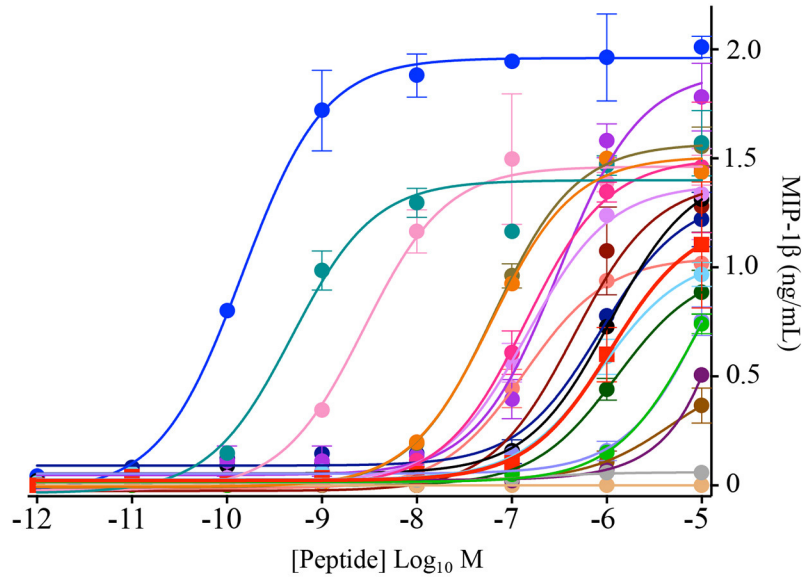
FIGURE 6 | Recognition of 72 randomly chosen and uniformly distributed bacterial peptides by the InsB4 T-cell clone. Each panel shows results of MIP-1β assay, indicating activation of T-cell clone InsB4 by 12 peptides and the index HLVEALYLV peptide (shown as a bold red line). InsB4 was incubated overnight with the indicated peptides and MIP-1β released into the supernatants quantified by ELISA. Bins 1-5 were performed in the same assay with Bin 6 performed in a separate experiment hence the different EC₅₀ values for the index peptide. EC₅₀ for each peptide (M) is displayed in the tables. NR, no T-cell response. The mean pEC₅₀ for each bin is plotted in **Figure 7A**.



a list of potential agonists by assigning to each peptide an agonist likelihood score (17). Testing with 144 peptides selected uniformly across a logarithmic scale of the lambda score from the fungal and bacterial pathoproteome with values spanning ~ 6 orders of magnitude demonstrated that the CPL-driven database searching could accurately identify peptides derived from cellular organisms that were recognised by an autoimmune T-cell clone. Functional testing of 20 high scoring peptides from the new bacterial and fungal databases showed that 20/20 and 18/20 acted as agonists of the InsB4 insulin-specific T-cell clone. Furthermore, 7/20 and 12/20 sequences predicted from the bacterial and fungal databases were more potent agonists than the HLVEALYLV index sequence. In theory, if presented *in vivo*, such super-agonist peptides might have the capacity to break tolerance to trigger disease. Importantly, one bacterial and one fungal peptide derived from *H. pylori* and *C. albicans*, respectively, were selected to demonstrate conclusively that they could engage the

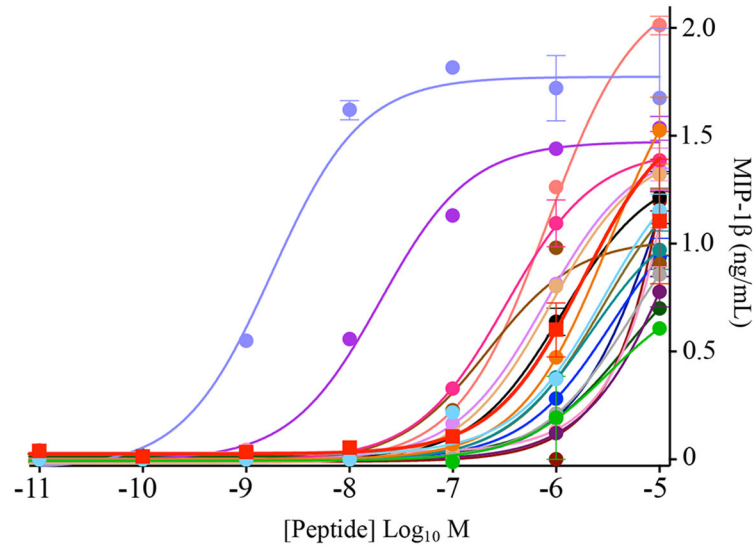
InsB4 autoimmune TCR as peptide-MHC tetramers and by SPR. Interestingly, several recent studies have shown an increased rate of *H. pylori* infection in T1D patients (42–45) leading to suggestions that it may be a disease trigger; however, it is also possible that T1D might render patients more susceptible to *H. pylori* infection. Similarly, T1D has also been linked with *C. albicans* at the time of diagnosis (46) but diabetes is known to predispose individuals to fungal infections (47) so the direction of any link is still in question.

The resemblance between epitopes derived from microbial and host proteins leading to crossreactivity of T-cells in the host has been termed molecular mimicry (48, 49) and is believed to be a major mechanism by which autoimmune diseases are triggered (50, 51). Indeed, T-cell crossreactivity between microbially-derived peptides and self-peptides can induce autoimmunity in experimental animal models (52–54). Recent studies have confirmed that TCRs are capable of recognising large numbers



Score	Peptide	EC ₅₀	Organism	Protein
-24.60	<u>HLVEALYLV</u>	1.1 x10 ⁻⁶	<i>Homo sapiens</i>	Insulin β chain
-13.48	<u>LILENYMFA</u>	8.8 x10⁻⁷	<i>Candida tropicalis</i>	Putative protein
-13.59	<u>IQLENSLFA</u>	8.9 x10 ⁻⁶	<i>Mucor circinelloides</i>	Putative protein
-13.61	<u>LLVERALFT</u>	NR	<i>Fusarium oxysporum</i>	Putative protein
-13.73	<u>MIVEGFLLL</u>	5.9 x10⁻⁸	<i>Sporothrix schenckii</i>	Putative protein
-13.81	<u>ITIERILFI</u>	8.5 x10 ⁻⁵	<i>Cryptococcus neoformans</i> <i>Rhodotorula glutinis</i> <i>Trichosporon asahii</i>	Putative protein Importin-alpha receptor Importin-alpha receptor
-13.82	<u>IMLENGPFL</u>	NR	<i>Aspergillus niger</i>	Putative protein
-13.84	<u>LIIENAPLI</u>	4.1 x10⁻¹⁰	<i>Saccharomyces cerevisiae</i>	Tae2p
-13.90	<u>MLVEGVLLA</u>	5.5 x10⁻⁸	<i>Mucor circinelloides</i>	Putative protein
-13.92	<u>MIIEGFLLI</u>	1.3 x10⁻⁷	<i>Mucor circinelloides</i>	Putative protein
-13.94	<u>LLVENWPLL</u>	2.4 x10⁻⁹	<i>Aspergillus niger</i>	Putative protein
-14.01	<u>LIIIEWIFFL</u>	1.5 x10⁻⁷	<i>Aspergillus clavatus</i>	Oligosaccharyltransferase
-14.14	<u>MQVEWILTF</u>	1.2 x10 ⁻⁵	<i>Candida albicans</i>	Putative protein
-14.19	<u>LQLEGGQLLL</u>	3.9 x10 ⁻⁶	<i>Fusarium oxysporum</i>	Putative protein
-14.22	<u>LVVEGSLFV</u>	3.4 x10 ⁻⁷	<i>Fusarium oxysporum</i> <i>Fusarium verticillioides</i>	Putative protein Putative protein
-14.24	<u>LIVEGILLT</u>	1.1 x10 ⁻⁶	<i>Aspergillus oryzae</i> <i>Aspergillus niger</i> <i>Penicillium expansum</i>	Putative protein Putative protein Putative protein
-14.31	<u>MIVENVPLL</u>	1.7 x10⁻¹⁰	<i>Candida albicans</i>	Putative protein
-14.32	<u>LLVEDMLLM</u>	3.9 x10⁻⁷	<i>Sporothrix schenckii</i>	Putative protein
-14.32	<u>LIIIEWYLIM</u>	8.6 x10⁻⁷	<i>Candida tropicalis</i>	Putative protein
-14.34	<u>MLLEGMMIV</u>	1.1 x10⁻⁷	<i>Fusarium oxysporum</i>	Putative protein
-14.34	<u>FLLERILIV</u>	9.6 x10⁻⁷	<i>Saccharomyces cerevisiae</i>	Swf1p

FIGURE 8 | InsB4 recognizes fungal peptides with increased sensitivity relative to index insulin peptide. The 20 top-scoring peptides from fungal species were tested overnight for activation on InsB4 by MIP-1β ELISA. The likelihood scores and EC₅₀ of peptide (M) activation are shown. Putative proteins are the product of predicted open reading frames from the genome sequence of the species indicated. NR, no T-cell response. Bold EC₅₀ more potent than the index peptide. Amino acid residues that are the same for each fungal peptide and the insulin index peptide (HLVEALYLV) are underlined.

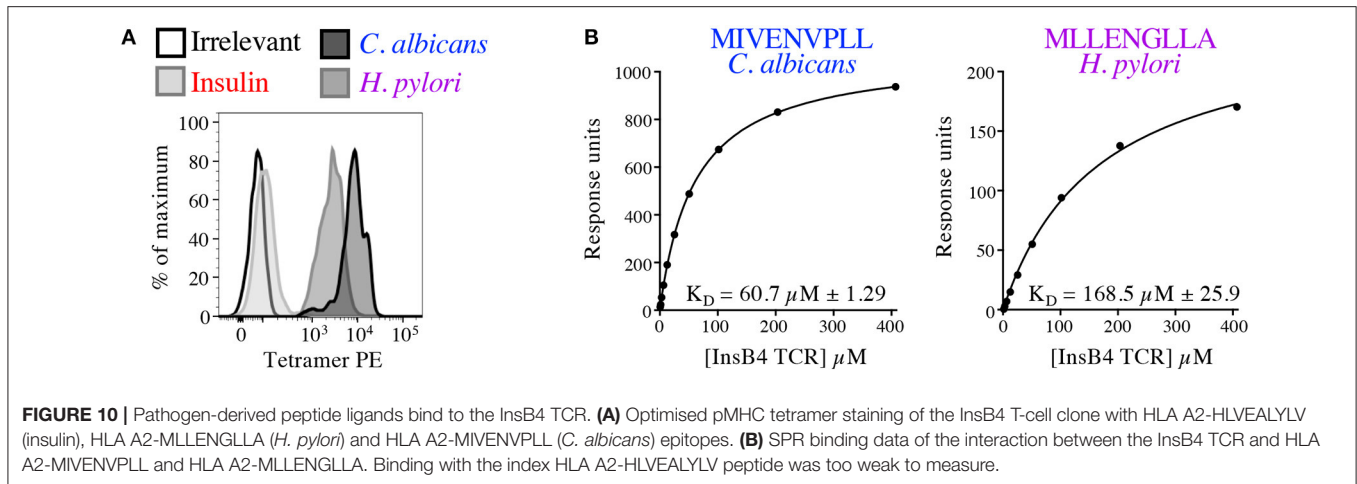


Score	Peptide	EC ₅₀	Organism	Protein
-24.60	<u>HLVEALYLV</u>	1.1 x10 ⁻⁶	<i>Homo sapiens</i>	Insulin β chain
-12.66	<u>LLIERYLLV</u>	2.9 x10 ⁻⁶	<i>Klebsiella pneumoniae</i>	Putative protein
			<i>Enterobacter cloacae</i>	Putative protein
-12.70	<u>IMIENILFF</u>	3.2 x10 ⁻⁶	<i>Bacillus cereus</i>	Putative protein
-12.83	<u>LLIERFLFL</u>	5.1 x10 ⁻⁶	<i>Helicobacter pylori</i>	Putative protein
-12.87	<u>LMLERALLV</u>	3.3 x10 ⁻⁶	<i>Helicobacter pylori</i>	Cytochrome c biogenesis protein
-12.88	<u>ILIERILLA</u>	1.5 x10 ⁻⁵	<i>Haemophilus influenzae</i>	Glycosyltransferase
-12.90	<u>IILENFLFF</u>	8.8 x10⁻⁷	<i>Streptococcus mitis</i>	Putative protein
-12.92	<u>LMLERALLF</u>	2.2 x10 ⁻⁶	<i>Helicobacter pylori</i>	Cytochrome c biogenesis protein
-12.93	<u>IILENILLF</u>	2.8 x10 ⁻⁶	<i>Clostridium novyi</i>	Hypothetical protein
-12.94	<u>LLIERGLLA</u>	3.2 x10⁻⁷	<i>Pseudomonas aeruginosa</i>	LysR family transcriptional regulator
			<i>Streptococcus oralis</i>	Putative membrane protein
-12.97	<u>LLVEGILLV</u>	4.6 x10 ⁻⁵	<i>Streptococcus pseudopneumoniae</i>	Putative membrane protein
-13.05	<u>MLVERYLLA</u>	7.8 x10⁻⁷	<i>Escherichia coli</i>	Putative prophage ATP/GTP binding protein
-13.07	<u>IIIEGILFV</u>	1.8 x10⁻⁹	<i>Streptococcus caballi</i>	Putative protein
-13.08	<u>ILIERFLLL</u>	2.2 x10⁻⁷	<i>Clostridium butyricum</i>	Putative protein
-13.13	<u>MLENGLLA</u>	1.9 x10⁻⁸	<i>Helicobacter pylori</i>	Putative protein
-13.14	<u>LMIERGLLA</u>	1.3 x10 ⁻⁶	<i>Pseudomonas aeruginosa</i>	Putative protein
-13.16	<u>LILERALLL</u>	3.5 x10 ⁻⁶	<i>Pseudomonas stutzeri</i>	Flagellar biosynthesis protein
			<i>Haemophilus aegyptius</i>	Putative protein
-13.17	<u>LIIENQLFF</u>	~0.08	<i>Haemophilus influenzae</i>	Putative protein
-13.24	<u>MLVENQLLV</u>	1.1 x10 ⁻⁵	<i>Pseudomonas mendocina</i>	AraC family transcriptional regulator
-13.26	<u>LIENALMF</u>	8.0 x10⁻⁷	<i>Pseudomonas stutzeri</i>	Polysaccharide biosynthesis protein
-13.28	<u>ILLERILIF</u>	4.2 x10 ⁻⁶	<i>Salmonella enterica</i>	Putative protein

FIGURE 9 | InsB4 recognizes bacterial peptides with increased sensitivity relative to index insulin peptide. The 20 top-scoring peptides from category II (i.e., confirmed pathogenic) bacterial species were tested overnight for activation on InsB4 by MIP-1β ELISA. The likelihood scores and EC₅₀ of peptide (M) activation are shown. Putative proteins are the product of predicted open reading frames from the genome sequence of the species indicated. Bold EC₅₀ more potent than the index peptide. ~ = ambiguous EC₅₀ based on GraphPad Prism software analysis. Amino acid residues that are the same for each bacterial peptide and the insulin index peptide (HLVEALYLV) are underlined.

of different individual peptides in the context of a single MHC molecule (7, 55, 56). Several mechanisms for TCR promiscuity have been suggested that include; macrolevel changes in peptide binding register, alterations in TCR crossing angle on

peptide-MHC (pMHC), TCR CDR loop flexibility, and residue focused “hotspot” binding (1). Evidence for these mechanisms is emerging and points to further unanticipated molecular flexibility that can extend the range of peptides recognised via



individual TCRs. The TCR can interact with different pMHCs using a common (55, 57–60), or a highly divergent binding mode (61). The structure of TCR CDR loops are known to differ in ligated and unligated TCR crystal structures suggesting that these loops are flexible and might be capable of binding to different peptides in different ways (59, 62). However, it is worth noting that a recent study found large conformational differences of up to 5 Å in individual CDR loops in different crystals of the same unligated TCR, suggesting that crystal structures may only offer an artefactual indication of TCR flexibility due to lattice packing and crystallisation conditions (63). Two recent studies have described extensive peptide crossreactivity as the result of constrained interactions with peptide hotspots in MHC class I and class II-restricted peptides (16, 55). More recent studies have shown that heretofore unanticipated rearrangements in the peptide and presenting MHC protein can extend the range of ligands that can be recognised (64, 65). The potential consequences of widespread T-cell crossreactivity are profound and range from heterologous immunity, where an individual T-cell clone can target more than one pathogen via a single TCR, to crossreactivity between pathogen-derived and self-peptides [reviewed in (1)].

While the combined computational-experimental approach presented in this study can be used to discover potential infectious triggers for autoimmune disease. Further work will be required to reveal if any of the potential infection-derived peptides found by our work are genuinely presented to T-cells *in vivo*. Moreover, this combined computational-experimental approach also has the potential to be used to discover potent peptide agonists of any T-cell clone for use in TCR-optimised peptide skewing of the repertoire of T-cells (TOPSORT) (66) and to rank the agonist potential of D-amino acid peptides (67). The GPU-accelerated strategy has not only surpassed the previous framework (17) in terms of speed, cost-, and power-efficiency, but it has also demonstrated its robustness by addressing the theoretical peptide ranking problem. In sum, GPU-based computing is practical and within reach of users with a desk-top computer; as such, we expect that it will find multiple applications

in other computationally demanding areas of immunology such as immune phenotyping and antigen receptor analysis.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Files**.

AUTHOR CONTRIBUTIONS

TW, GD, PB, AW, AF, LW, JH, MA, HB, MC, RK, DC, MP, AS, and BS performed and/or directed experiments, analysed data, and/or critiqued the manuscript. AS, LW, and BS conceived, funded, and directed the project. AS, TW, GD, and BS wrote the manuscript.

FUNDING

This work was funded by AS's Wellcome Senior Investigator Award (WT100327MA) and BS' Royal Society Research Grant (RSG\R1\180437) and used equipment and reagents purchased by UK Biotechnology and Biological Sciences Research Council (Grant BB/H001085/1). TW was a Ph.D. student funded by Cardiff University Systems Immunity Research Institute. AW was a Ph.D. student part funded by the Welsh Government via its National Research Network.

ACKNOWLEDGMENTS

The authors would like to thank Thomas Green of Advanced Research Computing at Cardiff (ARCCA) for his advice on profiling and benchmarking the GPU code.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00296/full#supplementary-material>

REFERENCES

- Sewell AK. Why must T cells be cross-reactive? *Nat Rev Immunol.* (2012) 12:669–77. doi: 10.1038/nri3279
- Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. *Science.* (1999) 286:958–61. doi: 10.1126/science.286.5441.958
- Attaf M, Huseby E, Sewell AK. alphabeta T cell receptors as predictors of health and disease. *Cell Mol Immunol.* (2015) 12:391–9. doi: 10.1038/cmi.2014.134
- Bridgeman JS, Sewell AK, Miles JJ, Price DA, Cole DK. Structural and biophysical determinants of alphabeta T-cell antigen recognition. *Immunology.* (2012) 135:9–18. doi: 10.1111/j.1365-2567.2011.03515.x
- Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature.* (1996) 384:134–41. doi: 10.1038/384134a0
- Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol Today.* (1998) 19:395–404. doi: 10.1016/S0167-5699(98)01299-7
- Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, et al. A single autoimmune T cell receptor recognizes more than a million different peptides. *J Biol Chem.* (2012) 287:1168–77. doi: 10.1074/jbc.M111.289488
- Nejentsev S, Howson JM, Walker NM, Szeszeko J, Field SF, Stevens HE, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature.* (2007) 450:887–92. doi: 10.1038/nature06406
- Bottazzo GF, Dean BM, McNally JM, MacKay EH, Swift PG, Gamble DR. *In situ* characterization of autoimmune phenomena and expression of HLA molecules in the pancreas in diabetic insulinitis. *N Engl J Med.* (1985) 313:353–60. doi: 10.1056/NEJM198508083130604
- DiLorenzo TP, Serreze DV. The good turned ugly: immunopathogenic basis for diabetogenic CD8+ T cells in NOD mice. *Immunol Rev.* (2005) 204:250–63. doi: 10.1111/j.0105-2896.2005.00244.x
- Marron MP, Graser RT, Chapman HD, Serreze DV. Functional evidence for the mediation of diabetogenic T cell responses by HLA-A2.1 MHC class I molecules through transgenic expression in NOD mice. *Proc Natl Acad Sci USA.* (2002) 99:13753–8. doi: 10.1073/pnas.212221199
- Bulek AM, Cole DK, Skowera A, Dolton G, Gras S, Madura F, et al. Structural basis for the killing of human beta cells by CD8(+) T cells in type 1 diabetes. *Nat Immunol.* (2012) 13:283–9. doi: 10.1038/ni.2206
- Kronenberg D, Knight RR, Estorninho M, Ellis RJ, Kester MG, de Ru A, et al. Circulating preproinsulin signal peptide-specific CD8 T cells restricted by the susceptibility molecule HLA-A24 are expanded at onset of type 1 diabetes and kill beta-cells. *Diabetes.* (2012) 61:1752–9. doi: 10.2337/db11-1520
- Skowera A, Ellis RJ, Varela-Calvino R, Arif S, Huang GC, Van-Krinks C, et al. CTLs are targeted to kill beta cells in patients with type 1 diabetes through recognition of a glucose-regulated preproinsulin epitope. *J Clin Invest.* (2008) 118:3390–402. doi: 10.1172/JCI35449
- Coppieters KT, Dotta F, Amirian N, Campbell PD, Kay TW, Atkinson MA, et al. Demonstration of islet-autoreactive CD8 T cells in insulinitic lesions from recent onset and long-term type 1 diabetes patients. *J Exp Med.* (2012) 209:51–60. doi: 10.1084/jem.20111187
- Cole DK, Bulek AM, Dolton G, Schauenberg AJ, Szomolay B, Rittase W, et al. Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *J Clin Invest.* (2016) 126:2191–204. doi: 10.1172/JCI85679
- Szomolay B, Liu J, Brown PE, Miles JJ, Clement M, Llewellyn-Lacey S, et al. Identification of human viral protein-derived ligands recognized by individual major histocompatibility complex class I (MHCI)-restricted T-cell receptors. *Immunol Cell Biol.* (2016) 4:573–82. doi: 10.1038/icb.2016.12
- Klus P, Lam S, Lyberg D, Cheung MS, Pullan G, McFarlane I, et al. BarraCUDA - a fast short read sequence aligner using graphics processing units. *BMC Res Notes.* (2012) 5:27. doi: 10.1186/1756-0500-5-27
- Lei G, Dou Y, Wan W, Xia F, Li R, Ma M, Zou D, et al. CPU-GPU hybrid accelerating the Zuker algorithm for RNA secondary structure prediction applications. *BMC Genomics.* (2012) 13(Suppl. 1):S14. doi: 10.1186/1471-2164-13-S1-S14
- Mashimo T, Fukunishi Y, Kamiya N, Takano Y, Fukuda I, Nakamura H. Molecular Dynamics simulations accelerated by GPU for biological macromolecules with a non-ewald scheme for electrostatic interactions. *J Chem Theory Comput.* (2013) 9:5599–609. doi: 10.1021/ct400342e
- Zhou Y, Liepe J, Sheng X, Stumpf MP, Barnes C. GPU accelerated biochemical network simulation. *Bioinformatics.* (2011) 27:874–6. doi: 10.1093/bioinformatics/btr015
- Nickols J, Buck I, Garland M, Skadron K. Scalable parallel programming with CUDA. *Queue.* (2008) 6:40–53. doi: 10.1145/1365490.1365500
- Tungatt K, Bianchi V, Crowther MD, Powell WE, Schauenberg AJ, Trimby A, et al. Antibody stabilization of peptide-MHC multimers reveals functional T cells bearing extremely low-affinity TCRs. *J Immunol.* (2015) 194:463–74. doi: 10.4049/jimmunol.1401785
- Flynn MJ. Some computer organizations and their effectiveness. *IEEE Trans Comput.* (1972) 948–60. doi: 10.1109/TC.1972.5009071
- Chang L, Kjer-Nielsen L, Flynn S, Brooks AG, Mannering SI, Honeyman MC, et al. Novel strategy for identification of candidate cytotoxic T-cell epitopes from human preproinsulin. *Tissue Antigens.* (2003) 62:408–17. doi: 10.1034/j.1399-0039.2003.00122.x
- Theaker SM, Rius C, Greenshields-Watson A, Lloyd A, Trimby A, Fuller A, et al. T-cell libraries allow simple parallel generation of multiple peptide-specific human T-cell clones. *J Immunol Methods.* (2016) 430:43–50. doi: 10.1016/j.jim.2016.01.014
- Knight RR, Dolton G, Kronenberg-Versteeg D, Eichmann M, Zhao M, Huang GC, et al. A distinct immunogenic region of glutamic acid decarboxylase 65 is naturally processed and presented by human islet cells to cytotoxic CD8 T cells. *Clin Exp Immunol.* (2015) 179:100–7. doi: 10.1111/cei.12436
- Purbhoo MA, Boulter JM, Price DA, Vuidepot AL, Hourigan CS, Dunbar PR, et al. The human CD8 coreceptor effects cytotoxic T cell activation and antigen sensitivity primarily by mediating complete phosphorylation of the T cell receptor zeta chain. *J Biol Chem.* (2001) 276:32786–92. doi: 10.1074/jbc.M102498200
- Laugel B, van den Berg HA, Gostick E, Cole DK, Wooldridge L, Boulter J, et al. Different T cell receptor affinity thresholds and CD8 coreceptor dependence govern cytotoxic T lymphocyte activation and tetramer binding properties. *J Biol Chem.* (2007) 282:23799–810. doi: 10.1074/jbc.M700976200
- Ekeruche-Makinde J, Miles JJ, van den Berg HA, Skowera A, Cole DK, Dolton G, et al. Peptide length determines the outcome of TCR/peptide-MHCI engagement. *Blood.* (2013) 121:1112–23. doi: 10.1182/blood-2012-06-437202
- Laugel B, Price DA, Milicic A, Sewell AK. CD8 exerts differential effects on the deployment of cytotoxic T lymphocyte effector functions. *Eur J Immunol.* (2007) 37:905–13. doi: 10.1002/eji.200636718
- Price DA, Sewell AK, Dong T, Tan R, Goulder PJ, Rowland-Jones SL, et al. Antigen-specific release of beta-chemokines by anti-HIV-1 cytotoxic T lymphocytes. *Curr Biol.* (1998) 8:355–8. doi: 10.1016/S0960-9822(98)70138-1
- Rius C, Attaf M, Tungatt K, Bianchi V, Legut M, Bovay A, et al. Peptide-MHC class I tetramers can fail to detect relevant functional T cell clonotypes and underestimate antigen-reactive T cell populations. *J Immunol.* (2018) 200:2263–79. doi: 10.4049/jimmunol.1700242
- Laugel B, Boulter JM, Lissin N, Vuidepot A, Li Y, Gostick E, et al. Design of soluble recombinant T cell receptors for antigen targeting and T cell inhibition. *J Biol Chem.* (2005) 280:1882–92. doi: 10.1074/jbc.M409427200
- Dolton G, Zervoudi E, Ruis C, Wall A, Thomas HL, Fuller A, et al. Optimised peptide-MHC multimer protocols for detection and isolation of autoimmune T-cells. *Front Immunol.* (2018) 9:1378. doi: 10.3389/fimmu.2018.01378
- Lissina A, Ladell K, Skowera A, Clement M, Edwards E, Seggewiss R, et al. Protein kinase inhibitors substantially improve the physical detection of T-cells with peptide-MHC tetramers. *J Immunol Methods.* (2009) 340:11–24. doi: 10.1016/j.jim.2008.09.014
- Cole DK, Dunn SM, Sami M, Boulter JM, Jakobsen BK, Sewell AK. T cell receptor engagement of peptide-major histocompatibility complex class I does not modify CD8 binding. *Mol Immunol.* (2008) 45:2700–9. doi: 10.1016/j.molimm.2007.12.009
- Cole DK, Fuller A, Dolton G, Zervoudi E, Legut M, Miles K, et al. Dual molecular mechanisms govern escape at immunodominant HLA A2-restricted HIV epitope. *Front Immunol.* (2017) 8:1503. doi: 10.3389/fimmu.2017.01503

39. Gostick E, Cole DK, Hutchinson SL, Wooldridge L, Tafuro S, Laugel B, et al. Functional and biophysical characterization of an HLA-A*6801-restricted HIV-specific T cell receptor. *Eur J Immunol.* (2007) 37:479–86. doi: 10.1002/eji.200636243
40. Bell N, Hoberock J. Thrust: A productivity-oriented library for CUDA. *GPU Comput Gems Jade Ed.* (2012) 26:359–71. doi: 10.1016/B978-0-12-385963-1.00026-5
41. Dolton G, Tungatt K, Lloyd A, Bianchi V, Theaker SM, Trimby A, et al. More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. *Immunology.* (2015) 146:11–22. doi: 10.1111/imm.12499
42. Ojetti V, Pitocco D, Bartolozzi F, Danese S, Migneco A, Lupascu A, et al. High rate of helicobacter pylori re-infection in patients affected by type 1 diabetes. *Diabetes Care.* (2002) 25:1485. doi: 10.2337/diacare.25.8.1485
43. Osman SM, Mubarak SM, Omer IM, Abdullah MA. Helicobacter pylori infection and the onset of type 1 diabetes mellitus in Sudanese children. *Sudan J Paediatr.* (2016) 16:59–66.
44. Salardi S, Cacciari E, Menegatti M, Landi F, Mazzanti L, Stella FA, et al. Helicobacter pylori and type 1 diabetes mellitus in children. *J Pediatr Gastroenterol Nutr.* (1999) 28:307–9. doi: 10.1097/00005176-199903000-00017
45. Zekry OA, Abd Elwahid HA. The association between Helicobacter pylori infection, type 1 diabetes mellitus, and autoimmune thyroiditis. *J Egypt Public Health Assoc.* (2013) 88:143–7. doi: 10.1097/01.EPX.0000437621.23560.de
46. Gursoy S, Kockar T, Atik SU, Onal Z, Onal H, Adal E. Autoimmunity and intestinal colonization by Candida albicans in patients with type 1 diabetes at the time of the diagnosis. *Korean J Pediatr.* (2018) 61:217–20. doi: 10.3345/kjp.2018.61.7.217
47. Rodrigues CF, Rodrigues ME, Henriques M. Candida sp. infections in patients with diabetes mellitus. *J Clin Med.* (2019) 8:E76. doi: 10.3390/jcm8010076
48. Oldstone MB. Molecular mimicry and immune-mediated diseases. *FASEB J.* (1998) 12:1255–65. doi: 10.1096/fasebj.12.13.1255
49. Oldstone MB. Molecular mimicry, microbial infection, and autoimmune disease: evolution of the concept. *Curr Top Microbiol Immunol.* (2005) 296:1–17. doi: 10.1007/3-540-30791-5_1
50. Albert LJ, Inman RD. Molecular mimicry and autoimmunity. *N Engl J Med.* (1999) 341:2068–74. doi: 10.1056/NEJM199912303412707
51. Cusick MF, Libbey JE, Fujinami RS. Molecular mimicry as a mechanism of autoimmune disease. *Clin Rev Allergy Immunol.* (2012) 42:102–11. doi: 10.1007/s12016-011-8294-7
52. Hemmer B, Fleckenstein BT, Vergelli M, Jung G, McFarland H, Martin R, et al. Identification of high potency microbial and self ligands for a human autoreactive class II-restricted T cell clone. *J Exp Med.* (1997) 185:1651–9. doi: 10.1084/jem.185.9.1651
53. Olson JK, Croxford JL, Calenoff MA, Dal Canto MC, Miller SD. A virus-induced molecular mimicry model of multiple sclerosis. *J Clin Invest.* (2001) 108:311–8. doi: 10.1172/JCI200113032
54. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell.* (1995) 80:695–705. doi: 10.1016/0092-8674(95)90348-8
55. Birnbaum ME, Mendoza JL, Sethi DK, Dong S, Glanville J, Dobbins J, et al. Deconstructing the peptide-MHC specificity of T cell recognition. *Cell.* (2014) 157:1073–87. doi: 10.1016/j.cell.2014.03.047
56. Wilson DB, Wilson DH, Schroder K, Pinilla C, Blondelle S, Houghten RA, et al. Specificity and degeneracy of T cells. *Mol Immunol.* (2004) 40:1047–55. doi: 10.1016/j.molimm.2003.11.022
57. Adams JJ, Narayanan S, Liu B, Birnbaum ME, Kruse AC, Bowerman NA, et al. T cell receptor signaling is limited by docking geometry to peptide-major histocompatibility complex. *Immunity.* (2011) 35:681–93. doi: 10.1016/j.immuni.2011.09.013
58. Holland CJ, Rizkallah PJ, Vollers S, Calvo-Calle JM, Madura F, Fuller A, et al. Minimal conformational plasticity enables TCR cross-reactivity to different MHC class II heterodimers. *Sci Rep.* (2012) 2:629. doi: 10.1038/srep00629
59. Reiser JB, Darnault C, Gregoire C, Mosser T, Mazza G, Kearney A, et al. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol.* (2003) 4:241–7. doi: 10.1038/ni891
60. Sethi DK, Gordo S, Schubert DA, Wucherpfennig KW. Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop. *Nat Commun.* (2013) 4:2623. doi: 10.1038/ncomms3623
61. Colf LA, Bankovich AJ, Hanick NA, Bowerman NA, Jones LL, Kranz DM, et al. How a single T cell receptor recognizes both self and foreign MHC. *Cell.* (2007) 129:135–46. doi: 10.1016/j.cell.2007.01.048
62. Tynan FE, Reid HH, Kjer-Nielsen L, Miles JJ, Wilce MC, Kostenko L, et al. A T cell receptor flattens a bulged antigenic peptide presented by a major histocompatibility complex class I molecule. *Nat Immunol.* (2007) 8:268–76. doi: 10.1038/ni1432
63. Holland CJ, MacLachlan BJ, Bianchi V, Hesketh SJ, Morgan R, Vickery O, et al. *In silico* and structural analyses demonstrate that intrinsic protein motions guide T cell receptor complementarity determining region loop flexibility. *Front Immunol.* (2018) 9:674. doi: 10.3389/fimmu.2018.00674
64. Madura F, Rizkallah PJ, Legut M, Holland CJ, Fuller A, Bulek A, et al. TCR-induced alteration of primary MHC peptide anchor residue. *Eur J Immunol.* (2019) 49:1052–66. doi: 10.1002/eji.201948085
65. Riley TP, Hellman LM, Gee MH, Mendoza JL, Alonso JA, Foley KC, et al. T cell receptor cross-reactivity expanded by dramatic peptide-MHC adaptability. *Nat Chem Biol.* (2018) 14:934–42. doi: 10.1038/s41589-018-0130-4
66. Galloway SAE, Dolton G, Attaf M, Wall A, Fuller A, Rius C, et al. Peptide super-agonist enhances T-cell responses to melanoma. *Front Immunol.* (2019) 10:319. doi: 10.3389/fimmu.2019.00319
67. Miles JJ, Tan MP, Dolton G, Edwards ES, Galloway SA, Laugel B, et al. Peptide mimic for influenza vaccination using nonnatural combinatorial chemistry. *J Clin Invest.* (2018) 128:1569–80. doi: 10.1172/JCI91512

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Whalley, Dolton, Brown, Wall, Wooldridge, van den Berg, Fuller, Hopkins, Crowther, Attaf, Knight, Cole, Peakman, Sewell and Szomolay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.