Author/s:
Mannering, SI;So, M;Elso, CM;Kay, TWH

Title:
Shuffling peptides to create T-cell epitopes: Does the immune system play cards

Date:
2018-01-01

Citation:
Mannering, S. I., So, M., Elso, C. M. & Kay, T. W. H. (2018). Shuffling peptides to create T-cell epitopes: Does the immune system play cards. Immunology and Cell Biology, 96 (1), pp.34-40. https://doi.org/10.1111/imcb.1015.

Persistent Link:
https://hdl.handle.net/11343/293995

Article type : Review

Shuffling peptides to create T-cell epitopes-does the immune system play cards?

Stuart I. Mannering, Michelle So, Colleen M. Elso and Thomas W.H. Kay

Immunology and Diabetes Unit, St. Vincent's Institute of Medical Research, 9 Princes St, Fitzroy, Victoria 3065, Australia

Running title: T-cell epitopes formed by protein splicing

Corresponding Author Stuart I. Mannering
Address: 9 Princes St, Fitzroy, Victoria 3065, Australia
Telephone: 61-3-9288-2480
Fax: 61-3-9416-2676
Email smannering@svi.edu.au

Conflict of interest statement: The authors have no conflicts of interest to declare.

Word count: 3,205

Figure and Table count: 1 figure, 2 tables

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/imcb.1015

Abstract

For a long time, immunologists have believed that classical $CD4^+$ and $CD8^+$ T cells recognize peptides (referred to as epitopes), derived from protein antigens presented by MHC/HLA class I or II. Over the past 10-15 years it has become clear that epitopes recognized by $CD8^+$, and more recently $CD4^+$ T cells, can be formed by protein splicing. Here we review the discovery of spliced epitopes recognized by tumour-specific human $CD8^+$ T cells. We discuss how these epitopes are formed and some of the unusual variants that have been reported. Now, over a decade since the first report, evidence is emerging that spliced $CD8^+$ T-cell epitopes are much more common, and potentially much more important, than previously imagined. Recent work has shown that epitopes recognized by $CD4^+$ T cells can also be formed by protein splicing. We discuss the recent discovery of spliced $CD4^+$ T-cell epitopes and their potential role as targets of autoimmune T-cell responses. Finally, we highlight some of the new questions raised from our growing appreciation of T-cell epitopes formed by peptide splicing.

Introduction

Immunology textbooks tell us that the primary function of the immune system is to distinguish between self and non-self and then destroy anything that is both non-self and dangerous. The adaptive T-cell arm of the immune system does this by recognizing 8-14 amino acid long linear peptides presented by HLA (or MHC in the mouse) class I, or class II [1, 2]. Implicit in this model is this notion that antigen processing and presentation cannot alter the sequence of amino acids because to do so would be to blur the line between self and non-self. As discussed below, it is now clear that processing of antigens for presentation by HLA class I does shuffle the sequence of some peptide fragments. More recently, it has been suggested that epitopes recognized by $CD4^+$ T cells may also be formed by fusing peptide fragments.

Both $CD4^+$ and $CD8^+$ T cells use their T-cell receptors (TCRs) to recognize peptide antigens presented by HLA class II and class I, respectively. $CD8^+$ T-cell epitopes derive from proteins within the cell which are digested by the proteasome, trimmed and loaded onto HLA class I before being transported to the cell's surface for scrutiny by $CD8^+$ T cells [3-5]. Epitopes

recognized by $CD4^+$ T cells derive from exogenous antigens that are taken up by professional antigen presenting cells where they are degraded by lysosomal proteases [4, 5]. Protein fragments are then loaded onto class II before the complex is transported to the cell surface for scrutiny by $CD4^+$ T cells.

Proteases catalyse the cleavage of peptide bonds, however, under the appropriate conditions they can also catalyse the formation of peptide bonds. Protease-mediated peptide bond formation has been observed in bacteria, plants, mammals and *in vitro*. For example, bacterial sortases catalyse the cleavage and re-formation of peptide bonds by transpeptidation to help create the cross-linked structure of bacterial cell walls [6] (reviewed by Patterson [7]. One of the earliest reported examples of protease-mediated protein splicing in eukaryotes was the proteolytic processing of the lectin, ConA, by the Jackbean (*Canavalia ensiformis)* [8]. The precursor protein is cleaved and 're-ligated' by an asparagine protease. 'Reverse proteolysis' of several proteins and peptides has also been observed *in vitro* [9-12].

Recently immunologists have begun to appreciate that reverse proteolysis can lead to the formation of T-cell epitopes. Our goal here is to review the discovery and the reports describing the analysis of T-cell epitopes formed by protease-mediated transpeptidation, or protein splicing. This review is divided into three sections: (i) a description of the discovery of spliced $CD8^+$ T-cell epitopes, (ii) a discussion of the mechanism of their formation, and (iii) a description of spliced protein epitopes recognized by $CD4^+$ T cells. Although still in its infancy, the analysis of spliced T-cell epitopes promises to teach us more about the exquisite subtlety of the immune system.

Spliced Class I epitopes

The first reported $CD8^+$ T-cell epitope formed by protein splicing were identified by the analysis of a tumour-specific $CD8^+$ T-cell clone in January 2004 (see Table 1). Hanada et al [13] followed the approach, pioneered by Boon et al [14], of isolating tumour-specific $CD8^+$ T-cell clones, determining their HLA restriction and then transfecting COS cells with plasmids encoding the restricting HLA and pools of a tumour-derived cDNA. In this case, the $CD8^+$ T-cell clone, C2, recognized an epitope derived from Fibroblast Growth Factor 5 (FGF-5)

presented by HLA-A3. Truncating the cDNA used to sensitise transfected COS cells for recognition by $CD8^+$ T cells narrowed the region encoding the epitope down to 179 base pairs (~60 amino acids), but surprisingly, shorter constructs did not activate the T-cell clone. By screening a large panel of cDNA constructs, with systematic deletions of the minimum sensitizing construct, it became clear that the epitope required amino acids more than 40 amino acids apart. Finally, the identity of the minimum stimulating 9mer peptide epitope was shown to be formed by splicing together two FGF-5-derived peptide fragments, NTYAS with LPRFK, and splicing out the intervening 40 amino acids. This created the epitope NTYAS-LPRFK which potently stimulated the clone C2.

Three months later, Vigneron et al reported a melanoma-derived epitope also formed by proteasome-mediated peptide splicing of $gp100^{PMEL17}$ [15]. They also took a functional approach using an HLA-A32 restricted, $gp100^{PMEL17}$ specific, $CD8^+$ T-cell clone called 'CTL 14'. In this case, the epitope could be narrowed down to a 13mer peptide (RTKAWNRQLYPEW) that required further processing before it could sensitise $HLA\text{-}A32^+$ APC for recognition by CTL 14. The minimum epitope was found to be RTK-QLYPEW (Table 1) [16].

One corollary of the proteasome-mediated transpeptidation hypothesis is that peptide fragments could re-form peptide bonds in the opposite order relative to the native protein. Warren et al [17] reported the first 'retro-spliced' $CD8^+$ T-cell epitope in 2006. Their analysis of a $CD8^+$ T-cell clone specific for a minor histocompatibility antigen, SP110, led them to identify a 20 amino acid fragment of SP110 that was the minimum that stimulated the SP110 specific clone (CTL DRN-7). No synthetic peptides, including spliced peptides, could sensitise the EBV-transformed B cells used as APC for recognition by the $CD8^+$ T-cell clone. However, they found that two synthetic 10mers could stimulate recognition, but only when they were electroporated into the EBV cells at the same time. This led them to test synthetic peptides that mimicked the sequence of the SP110 epitope spliced in reverse order which they found could stimulate the clone (Tab1e 1). In addition to describing the first 'retro-spliced' $CD8^+$ T-cell epitope Warren et al also confirmed that the 'retro-spliced' epitope generation was also dependent upon proteasome processing [17].

Spliced epitopes may incorporate other posttranslational modifications (Table 1). Dalet et al [18], described a second 'retro-spliced' CD8 epitope. This HLA-A24 restricted epitope is derived from tyrosinase and identified by screening a $CD8^+$ T-cell clone. In addition to being retro-spliced this epitope comprises two asparagine residues which are deamidated to form aspartic acid residues [18]. Hence this epitope is formed by two posttranslational modifications: protein splicing and deamidation of two asparagine residues.

Peptide splicing can introduce an iso-peptide bond between the two protein fragments [19]. The peptide bonds between adjacent amino acids in proteins and peptides form between the α amino group of the C-terminal amino acid and the carboxylic acid of the N-terminal amino acid. When the C-terminal amino acid has an ε–amino group, as in lysine or arginine, an isopeptide bond may also form with the side chain, or ε–amino group. Using model synthetic peptides, Berkers et al [19] found that proteasome-mediated peptide splicing could generate spliced peptides that contained an isopeptide bond between the N-terminal lysine and the C-terminal peptide fragment. Interestingly, they estimated that spliced peptides with isopeptide bonds were formed one-tenth as often as spliced peptides with a standard peptide bond.

Mechanisms and motifs

As described above, spliced $CD8^+$ T-cell epitopes are formed by proteasome-mediated transpeptidation. Vigneron et al [15] showed *in vitro* that purified proteasomes generated a spliced epitope from $gp100^{MEL17}$. The epitope could be generated only when longer peptides, that required cleaving, were present. The epitope could not form when the two fragments (RTK and QLYPEW) that comprise the epitope, were delivered separately. Furthermore, the epitope formation was blocked by the proteasome inhibitor, lactacystin. The data led to the hypothesis that the catalytically active threonine residues in the standard proteasome form O-acyl intermediates with the N-terminal peptide fragment which is then subject to nucleophilic attack by the N-terminus of another peptide fragment, which forms the C-terminal end of the spliced peptide (Figure 1). This mechanism has since been supported by several other studies [15, 17-21]. In addition, Ebstein et al [21] have suggested that proteasome-mediated peptide splicing may also occur via a condensation reaction, within the barrel of the proteasome.

Mammalian cells can express two types of proteasome, the 'standard' proteasome or an 'immune' proteasome [22, 23]. Professional antigen presenting cells, such as DC, constitutively express the immune proteasome and other cells express it after exposure to IFNγ. Dalet et al [24] compared the efficiency of immune and standard proteasomes in generating spliced epitopes from FGF-5 [13], gp100^{PMEL17} [15] and SP110 [17]. The spliced epitopes derived from FGF-5 and gp100^{MEL17} were produced more efficiently by the standard proteasome, whereas the immune proteasome generated the SP110 derived epitope more efficiently [24]. The differences in epitope formation were attributed to different cleavage site preferences of the two types of proteasome [25]. Proteasomes that didn't favour the formation of the epitope cleaved the precursor protein in a position that destroyed the epitope [24]. Hence, transpeptidation is a common property of both types of proteasome, but the specificity of proteasome cleavage can bias towards or against a particular epitope.

Further investigation into the mechanism of peptide splicing has revealed some preference for the amino acids at the cleavage site [20]. The N-terminal peptide has the greatest impact on the rate of splicing. N-terminal fragments with aspartate, or polar uncharged residues at their C-terminus had a greater rate of splicing. In contrast, the residues in the C-terminal fragment had a minor impact on splicing efficiency. Perhaps not surprisingly, the concentration of the peptide fragment had the greatest impact on the efficiency of peptide splicing. More abundant fragments promote formation of spliced epitopes [20].

How common are spliced CD8$^+$ T-cell epitopes?

When they were first discovered, in 2004, epitopes formed by proteasome-mediated splicing were thought to be very rare; ~0.01 [15] to 0.0002% [13] of CD8 epitopes. The number of potential spliced peptides is enormous making it very challenging to search for these without any knowledge of their sequence. Until recently researchers could only search for epitopes that exactly matched the known protein sequences, but recent advances in bioinformatics now allow putative spliced epitopes to be identified more readily. Liepe et al [26] observed that a large proportion of peptides presented by HLA-A2 are formed by peptide splicing. When they analysed epitopes presented by B lymphoblastoid cell lines and primary human fibroblasts they found that spliced epitopes accounted for approximately one third

of the diversity and a quarter of the abundance of peptides bound to HLA-A2**.** This surprising observation is supported by *in vitro* proteasome-mediated digestion experiments [20]. Ebstein et al [21], for example, reported that spliced epitopes derived from the melanoma antigen $gp100^{mel}$ were generated as efficiently as unspliced epitopes. Although this abundance of *in vivo* spliced epitopes has yet to be confirmed by many other studies, it is remarkable that in little over 10 years, spliced CD8 T-cell epitopes have gone from curious oddities to being considered common constituents of the class I peptidome.

Class II epitopes formed by transpeptidation.

Peptide epitopes presented to $CD4^+$ T cells by class II MHC/HLA can also form by peptide splicing. Currently the only example of spliced $CD4^+$ T-cell epitopes comes from the study of the immunology of type 1 diabetes [27]. Type 1 diabetes is caused by an autoimmune T-cell response directed against the insulin-producing beta cells, which are found in the Islets of Langerhans (referred to as islets) in the pancreas. While it is clear that the disease is autoimmune [28] the antigens and epitopes recognized by pathogenic T cells are not well defined. The Non-Obese Diabetic (NOD) mouse is the most widely studied animal model of human type 1 diabetes. In both mice and humans autoimmune, or type 1, diabetes is characterized by the infiltration of the pancreatic islets by immune cells which destroy the insulin-producing beta cells [29, 30].

In 1989 Haskins et al [31] described a panel of $CD4^+$ T-cell clones (known by the prefix BDC) that precipitate the onset of diabetes when transferred to young NOD mice. Work with the BDC clones and TCR transgenic mouse lines derived from them confirmed that these T cells caused autoimmune diabetes in NOD mice [32, 33]. However, the beta-cell derived epitope(s) recognized by these clones could not be determined [34]. A breakthrough came in 2016 when Delong et al [27] reported that several of the BDC panel of T-cell clones recognized beta-cell derived peptides formed by protein splicing. They called these epitopes 'hybrid insulin peptides', or HIPs, because they all comprise a fragment of proinsulin fused to a fragment of another beta cell protein. The clones BDC-2.5, 9.46 and 10.1 recognize an epitope formed by the fusion of proinsulin's C-peptide to chromogranin A [27]. Similarly, the clones BDC-9.3 and 6.9 recognize an epitope formed by the fusion of C-peptide to another beta-cell protein, islet amyloid polypeptide-2 (IAPP2) (Table 2). Synthetic peptides, that mimic this sequence

219 are extremely potent stimulators of these clones, stimulating responses at low nanomolar
220 concentrations [27].
221
222 Human islet-infiltrating $CD4^+$ T cells, isolated from the residual pancreatic islets of deceased
223 organ donors who suffered from type 1 diabetes also recognize HIPs [27]. Two HIPs were
224 shown to stimulate human islet-infiltrating $CD4^+$ T cells; a Cpeptide-IAPP2; and a Cpeptide-
225 Neuropeptide-Y HIP [27]. Synthetic peptides of these sequences were also very also potent
226 stimulators of these T cells; responses were detected at low nanomolar concentrations. The
227 analysis of islet-infiltrating $CD4^+$ T cells from other deceased organ donors who suffered
228 from type 1 diabetes revealed two new HIPs [35]. These HIPs are formed by the fusion of
229 proinsulin C-peptide with fragments of islet amyloid polypeptide 1 (IAPP1) and insulin A
230 chain. In contrast to $CD8^+$ T-cell epitopes, most HIPs are formed in *trans*, when fragments
231 from two different proteins fuse. The exception is the C-peptide-insulin A chain HIP which is
232 derived entirely from proinsulin.
233
234 The discovery of $CD4^+$ T-cell epitopes formed by transpeptidation that are recognized by
235 human islet-infiltrating T cells strongly suggests that $CD4^+$ T-cell responses against HIPs may
236 play a direct pathogenic role in human T1D. This possibility is supported by the observation
237 that many HIPs are presented by HLA-DQ8 [27], which is well-known to greatly increase an
238 individual's risk of developing T1D [36, 37]. However, it is worth noting that islet-infiltrating
239 $CD4^+$ T cells also recognize a variety of 'regular' epitopes; most are derived from proinsulin
240 [35, 38, 39].
241
242 **How do HIPs form in beta cells?**
243 Currently, the pathway that leads to the formation of spliced CD4 T-cell epitopes in beta
244 cells is not clear. There is no evidence to suggest that the proteasome is responsible, instead
245 we suggest that formation of HIPs in beta cells arises from the active proteolysis that occurs
246 in normally functioning beta cells.
247
248 Insulin-secreting pancreatic beta cells contain granules in which insulin is stored prior to
249 release into the circulation. Beta-cell granules are the site of very active proteolysis. Insulin
250 is produced as a prohormone, proinsulin from which the central C-peptide is cleaved by

three enzymes, prohormone convertase (PC) 1, PC2 and carboxylpeptidase E [40]. After excision of the C-peptide, insulin crystalizes in the beta-cell granules. C-peptide remains in solution and is co-secreted with insulin [41]. In addition to proinsulin, many proteins, including chromogranins and islet amyloid poly peptides (IAPPs), are cleaved by the granule proteases to generate an array of bio-active peptides [42]. Under most circumstances 'reverse proteolysis' is very inefficient, but high protein concentrations in a confined environment favour protease mediated peptide fusion [43]. Because the beta-cell granules contain active proteases and they are densely packed with insulin, C-peptide and other proteins, they may be an ideal crucible for forming spliced peptides in a similar way to the proteasome. Hybrid insulin peptides are thought to be presented to islet-infiltrating $CD4^+$ T cells by macrophages and B cells that infiltrate human and NOD mouse islets [44, 45]. However, we have reported that murine and human beta cells can express MHC/HLA class II [46], so it is possible that hybrid insulin peptides may be directly presented to $CD4^+$ T cells by class II positive beta cells.

Concluding comments and outstanding questions

The observation that both $CD4^+$ and $CD8^+$ T-cell epitopes form by posttranslational splicing of self-proteins was unexpected. It is now clear that these epitopes form during protein digestion in the proteasome and that some $CD8^+$ T cells, within the normal repertoire, recognize these epitopes. Once thought of as a curiosity, epitopes formed by proteasome-mediated splicing may be very common [26]. Compared to the study of $CD8^+$ T-cell epitopes formed by peptide splicing, the study of $CD4^+$ T-cell epitopes is only just beginning. Remarkably, even at this very early stage the evidence suggests that these epitopes may play a role in a common autoimmune disease –type 1 diabetes.

The discovery of T-cell epitopes formed by peptide splicing raises questions about discrimination between self and non-self. The immune system uses both central and peripheral tolerance mechanisms to maintain tolerance to self antigens. Central tolerance refers to the purging from the TCR repertoire of T cells that express TCRs with inappropriate affinities to self HLA/peptide complexes within the thymus. The cortical thymic epithelial cells (cTECs) express a unique combination of proteasome subunits, the 'thymoproteasome' which facilitates positive selection [47]. Developing T cells migrate to the medulla where they

undergo negative selection mediated by medullary thymic epithelial cells (mTECs). mTECs express an array of tissue specific antigens driven by the transcriptional regulator AIRE [48]. Currently it is not known if thymic epithelial cells present the full array of spliced $CD8^+$ epitopes found in the periphery, or if T cells that recognize these epitopes are subject to negative selection. It is unlikely that HIPs, recognized by $CD4^+$ T cells will be found in the thymus because the protein processing machinery required is not found in thymic epithelial cells.

Peripheral tolerance refers to the induction of tolerance that occurs outside the thymus. Currently it is not clear how the formation of spliced peptide epitopes relates pheripheral tolerance. However, it is tempting to speculate that it is induced in a similar way to 'regular' antigens [49].

Other questions to be addressed include: has the proteasome pathway evolved to promote peptide splicing? At least it hasn't been selected against. Why is that? Are spliced epitopes the targets of T-cell responses against pathogen-derived proteins? Can we apply our emerging knowledge of T-cell specificity to develop $CD8^+$ T-cell therapies to treat cancer? Conversely, can HIPs be used to restore immune tolerance in type 1 diabetes and, if similar epitopes are involved, in other autoimmune diseases? If spliced peptides presented by both class I and class II are common, how is immunological tolerance maintained when the distinction between self and non-self becomes so blurred?

How can we address these questions? Advances in mass spectrometry and bioinformatics are driving a rapid increase in our knowledge of spliced epitopes. These approaches help us to identify the peptides that form, but don't give us any insight into the T-cell responses. New approaches to identifying the epitopes recognized by T cells are required to help screen the enormous array of possible peptide epitopes that may exist. When developed these methods supported by mass spectrometry and bioinformatics will no doubt reveal new levels of subtlety in the regulation of the immune system. It is clear that nature shuffles the deck of epitope cards – now our challenge is to work out the rules of the game.

References

1. Gotch F, Rothbard J, Howland K, Townsend A, McMichael A. Cytotoxic T lymphocytes recognize a fragment of influenza virus matrix protein in association with HLA-A2. *Nature* 1987; **326**: 881-882.

2. Babbitt BP, Allen PM, Matsueda G, Haber E, Unanue ER. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 1985; **317**: 359-361.

3. Rock KL, Reits E, Neefjes J. Present Yourself! By MHC Class I and MHC Class II Molecules. *Trends Immunol* 2016; **37**: 724-737.

4. Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature reviews. Immunology* 2011; **11**: 823-836.

5. Jensen PE. Recent advances in antigen processing and presentation. *Nature immunology* 2007; **8**: 1041-1048.

6. Marraffini LA, Dedent AC, Schneewind O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol Mol Biol Rev* 2006; **70**: 192-221.

7. Paterson GK, Mitchell TJ. The biology of Gram-positive sortase enzymes. *Trends Microbiol* 2004; **12**: 89-95.

8. Carrington DM, Auffret A, Hanke DE. Polypeptide ligation occurs during post-translational modification of concanavalin A. *Nature* 1985; **313**: 64-67.

9. Fodor S, Zhang Z. Rearrangement of terminal amino acid residues in peptides by protease-catalyzed intramolecular transpeptidation. *Anal Biochem* 2006; **356**: 282-290.

10. Hänsler M, Jakubke H-D. Nonconventional Protease Catalysis in Frozen Aqueous Solutions. *Journal of Peptide Science* 1996; **2**: 279-289.

11. Kumaran S, Datta D, Roy RP. Conformationally driven protease-catalyzed splicing of peptide segments: V8 protease-mediated synthesis of fragments derived from thermolysin and ribonuclease A. *Protein Sci* 1997; **6**: 2233-2241.

12. Kumaran S, Datta D, Roy RP. An enigmatic peptide ligation reaction: protease-catalyzed oligomerization of a native protein segment in neat aqueous solution. *Protein Sci* 2000; **9**: 734-741.

13. Hanada K, Yewdell JW, Yang JC. Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* 2004; **427**: 252-256.

14. Coulie PG, Van den Eynde BJ, van der Bruggen P, Boon T. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nat Rev Cancer* 2014; **14**: 135-146.

15. Vigneron N, Stroobant V, Chapiro J, Ooms A, Degiovanni G, Morel S *et al.* An antigenic peptide produced by peptide splicing in the proteasome. *Science* 2004; **304**: 587-590.

16. Cresswell P. Cell biology. Cutting and pasting antigenic peptides. *Science* 2004; **304**: 525-257.

17. Warren EH, Vigneron NJ, Gavin MA, Coulie PG, Stroobant V, Dalet A *et al.* An antigen produced by splicing of noncontiguous peptides in the reverse order. *Science* 2006; **313**: 1444-1447.

18. Dalet A, Robbins PF, Stroobant V, Vigneron N, Li YF, El-Gamil M *et al.* An antigenic peptide produced by reverse splicing and double asparagine deamidation.

Proceedings of the National Academy of Sciences of the United States of America 2011; **108**: E323-231.

19. Berkers CR, de Jong A, Schuurman KG, Linnemann C, Geenevasen JA, Schumacher TN *et al.* Peptide Splicing in the Proteasome Creates a Novel Type of Antigen with an Isopeptide Linkage. *Journal of immunology* 2015; **195**: 4075-4084.

20. Berkers CR, de Jong A, Schuurman KG, Linnemann C, Meiring HD, Janssen L *et al.* Definition of Proteasomal Peptide Splicing Rules for High-Efficiency Spliced Peptide Presentation by MHC Class I Molecules. *Journal of immunology* 2015; **195**: 4085-4095.

21. Ebstein F, Textoris-Taube K, Keller C, Golnik R, Vigneron N, Van den Eynde BJ *et al.* Proteasomes generate spliced epitopes by two different mechanisms and as efficiently as non-spliced epitopes. *Sci Rep* 2016; **6**: 24032.

22. Gaczynska M, Rock KL, Goldberg AL. Gamma-interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes. *Nature* 1993; **365**: 264-7.

23. Rock KL, Goldberg AL. Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annual review of immunology* 1999; **17**: 739-779.

24. Dalet A, Stroobant V, Vigneron N, Van den Eynde BJ. Differences in the production of spliced antigenic peptides by the standard proteasome and the immunoproteasome. *Eur J Immunol* 2011; **41**: 39-46.

25. Emmerich NP, Nussbaum AK, Stevanovic S, Priemer M, Toes RE, Rammensee HG *et al.* The human 26 S and 20 S proteasomes generate overlapping but different sets of peptide fragments from a model protein substrate. *The Journal of biological chemistry* 2000; **275**: 21140-21148.

26. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A *et al.* A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* 2016; **354**: 354-358.

27. Delong T, Wiles TA, Baker RL, Bradley B, Barbour G, Reisdorph R *et al.* Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science* 2016; **351**: 711-714.

28. Mannering SI, Pathiraja V, Kay TW. The case for an autoimmune aetiology of type 1 diabetes. *Clinical and experimental immunology* 2016; **183**: 8-15.

29. Hattori M, Buse JB, Jackson RA, Glimcher L, Dorf ME, Minami M *et al.* The NOD mouse: recessive diabetogenic gene in the major histocompatibility complex. *Science* 1986; **231**: 733-735.

30. In't Veld P. Insulitis in human type 1 diabetes: a comparison between patients and animal models. *Seminars in immunopathology* 2014; **36**: 569-579.

31. Haskins K, Portas M, Bergman B, Lafferty K, Bradley B. Pancreatic islet-specific T-cell clones from nonobese diabetic mice. *Proceedings of the National Academy of Sciences of the United States of America* 1989; **86**: 8000-8004.

32. Haskins K, McDuffie M. Acceleration of diabetes in young NOD mice with a CD4+ islet-specific T cell clone. *Science* 1990; **249**: 1433-1436.

33. Katz JD, Wang B, Haskins K, Benoist C, Mathis D. Following a diabetogenic T cell from genesis through pathogenesis. *Cell* 1993; **74**: 1089-1100.

34. Stadinski BD, Delong T, Reisdorph N, Reisdorph R, Powell RL, Armstrong M *et al.* Chromogranin A is an autoantigen in type 1 diabetes. *Nature immunology* 2010; **11**: 225-231.

35. Babon JA, DeNicola ME, Blodgett DM, Crevecoeur I, Buttrick TS, Maehr R *et al.* Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. *Nat Med* 2016; **22**: 1482-1487.

36. Todd JA, Bell JI, McDevitt HO. HLA-DQ beta gene contributes to susceptibility and resistance to insulin-dependent diabetes mellitus. *Nature* 1987; **329**: 599-604.

37. Noble JA. Immunogenetics of type 1 diabetes: A comprehensive review. *J Autoimmun* 2015; **64**: 101-112.

38. Pathiraja V, Kuehlich JP, Campbell PD, Krishnamurthy B, Loudovaris T, Coates PT *et al.* Proinsulin-Specific, HLA-DQ8, and HLA-DQ8-Transdimer-Restricted CD4+ T Cells Infiltrate Islets in Type 1 Diabetes. *Diabetes* 2015; **64**: 172-182.

39. Michels AW, Landry LG, McDaniel KA, Yu L, Campbell-Thompson M, Kwok WW *et al.* Islet-Derived CD4 T Cells Targeting Proinsulin in Human Autoimmune Diabetes. *Diabetes* 2017; **66**: 722-734.

40. Steiner DF, Park SY, Stoy J, Philipson LH, Bell GI. A brief perspective on insulin production. *Diabetes, obesity & metabolism* 2009; **11 Suppl 4:** 189-196.

41. Dodson G, Steiner D. The role of assembly in insulin's biosynthesis. *Curr Opin Struct Biol* 1998; **8**: 189-194.

42. Suckale J, Solimena M. The insulin secretory granule as a signaling hub. *Trends in Endocrinology & Metabolism* 2010; **21**: 599-609.

43. Somalinga BR, Roy RP. Volume exclusion effect as a driving force for reverse proteolysis. Implications for polypeptide assemblage in a macromolecular crowded milieu. *The Journal of biological chemistry* 2002; **277**: 43253-43261.

44. Carrero JA, Calderon B, Towfic F, Artyomov MN, Unanue ER. Defining the transcriptional and cellular landscape of type 1 diabetes in the NOD mouse. *PLoS One* 2013; **8**: e59701.

45. Willcox A, Richardson SJ, Bone AJ, Foulis AK, Morgan NG. Analysis of islet inflammation in human type 1 diabetes. *Clinical and experimental immunology* 2009; **155**: 173-181.

46. Zhao Y, Scott NA, Quah HS, Krishnamurthy B, Bond F, Loudovaris T *et al.* Mouse pancreatic beta cells express MHC class II and stimulate CD4+ T cells to proliferate. *European Journal of Immunology* 2015**; 45**: 2494-2503.

47. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y *et al.* Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science* 2007; **316**: 1349-1353.

48. Anderson MS, Su MA. AIRE expands: new roles in immune tolerance and beyond. *Nature reviews. Immunology* 2016; **16**: 247-258.

49. Lukacs-Kornek V, Turley SJ. Self-antigen presentation by dendritic cells and lymphoid stroma and its implications for autoimmunity. *Current opinion in immunology* 2011; **23**: 138-145.

50. Michaux A, Larrieu P, Stroobant V, Fonteneau JF, Jotereau F, Van den Eynde BJ *et al.* A spliced antigenic peptide comprising a single spliced amino acid is produced in the proteasome by reverse splicing of a longer peptide fragment followed by trimming. *Journal of immunology* 2014; **192**: 1962-1971.

Acknowledgements

SIM was supported by grants from: The Juvenile Diabetes Research Foundation (JDRF 17-2011-527), The Australian National Health and Medical Research Council (NHMRC GNT1123586), Diabetes Australia Research Trust Millennium Award T1D (Y17M1-MANS) and JDRF-Australia Clinical Research Network, Innovation Award. MS was supported by an NHMRC Postgraduate Scholarship (APP1094337) and a JDRF-Australia PhD Top-up Scholarship. TK is supported by an NHMRC Program Grant (APP1037321) and JDRF/Department of Health and Aging grants. All authors acknowledge the support from the Operational Infrastructure Support Program of the Victorian Government.

Table 1. $CD8^+$ T-cell epitopes formed by protein splicing defined by analysing T-cell specificity

Protein	Epitope	HLA restriction	Reference
FGF-5	NTYAS-PRFK	HLA-A3	13
$gp100^{PMEL17}$	RTK-QLYPEW	HLA-A32	15
SP110 (Minor H Ag)	SLPRGT-STPK*	HLA-A0301	17
Tyrosinase	IYM**D**GT-A<u>D</u>FSF*	HLA-A24	18
$gp100^{PMEL17}$	RSYVPLAH-R	HLA-A0301	50

*Reverse order relative to the native protein, <u>underlined</u> amino acids are formed by posttranslational modification

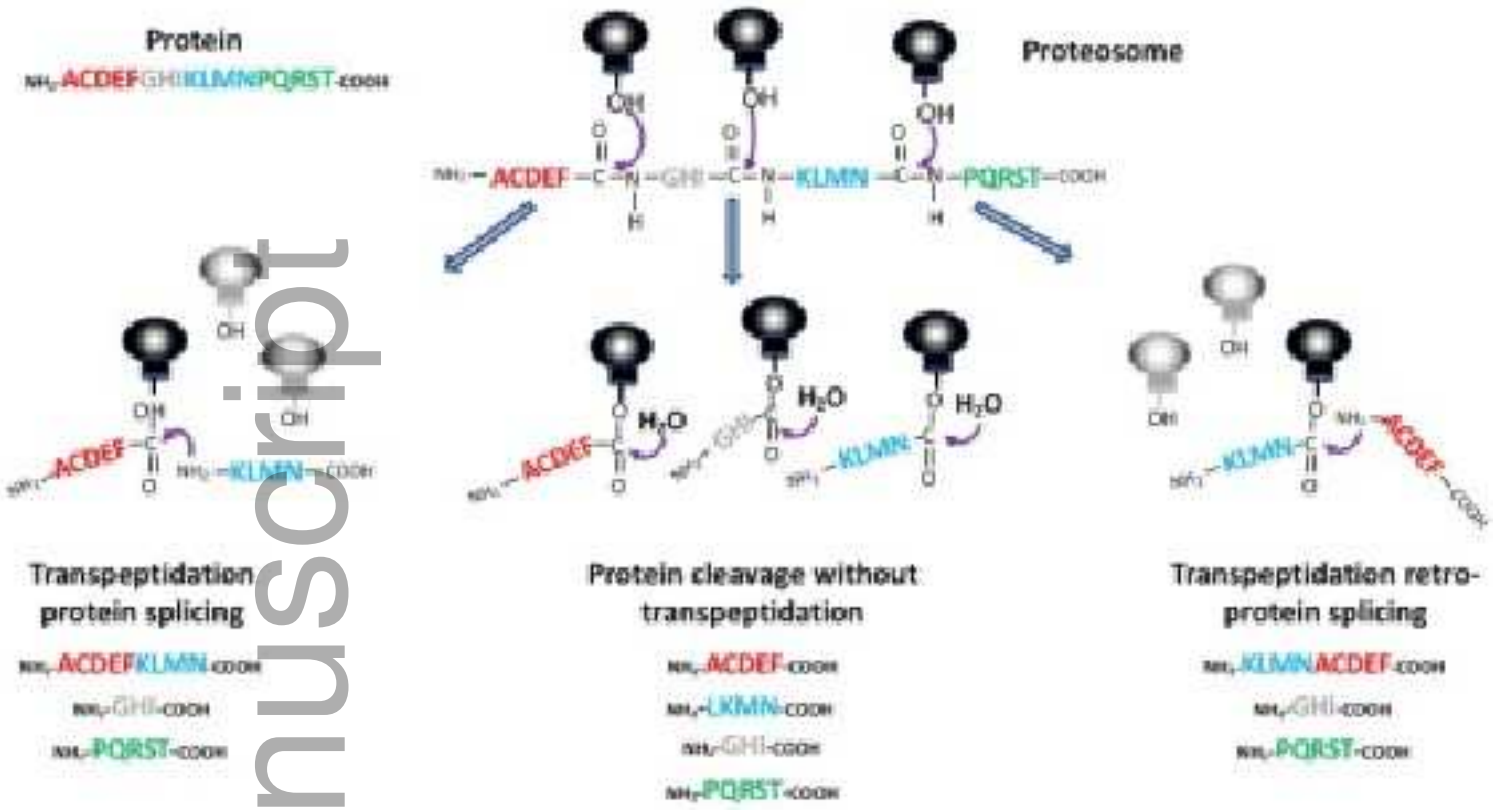
Table 2. Summary of $CD4^+$ T-cell epitopes formed by protein splicing defined by analysing T-cell specificity

Species	Clone	HIP	Restriction	Reference
NOD mouse	BDC 2.5	Cpept-ChgA (WE14)	$I\text{-}A^{g7}$	27
	BDC 9.46	Cpept-ChgA (WE14)	$I\text{-}A^{g7}$	
	BDC 10.1	Cpept-ChgA (WE14)	$I\text{-}A^{g7}$	
	BDC 9.3	Cpept-IAPP2	$I\text{-}A^{g7}$	
	BDC 6.9	Cpept-IAPP2	$I\text{-}A^{g7}$	
Human	A2.11	Cpept-IAPP2	HLA-DQ8	27

	A3.10	Cpept-IAPP2	HLA-DQ8	
	MG1*	Cpept-NeuroY	HLA-DQ8	
Human	$CD4^+$ T-cell line	Cpept-Insulin A chain	ND	35
	$CD4^+$ T-cell line	Cpept-IAPP1	ND	
	$CD4^+$ T-cell line	Cpept-IAPP2	ND	

Figure 1. Mechanism of proteasome-mediated peptide splicing.

This model shows the splicing of peptides within the proteasome. The putative peptide sequence is indicated by letters coloured to highlight the different fragments. The OH-groups of the catalytically active threonine are shown attached to the dark shapes which represent the wall of the proteasome. After peptide cleavage three outcomes are possible. From left to right; a spliced peptide may form by nucleophilic attack by another peptide fragment, the peptide fragment can be released by hydrolysis, or a spliced peptide with the fragments in the reverse order relative to the original protein may form. The putative peptide sequences, formed by each pathway, are shown below.



imcb_1015_f1.jpg