Beyond Degradation: How Reverse Proteolysis Creates Disease-Relevant Antigens

S. Yasin Tabatabaei Dakhili [1], Preety Panwar [1], Olivier Hinse [2], Jason Rogalski [2], Leonard J. Foster [2], Dieter Brömme[1,2]

Affiliations:

[1]Faculty of Dentistry, Department of Oral, Biological and Medical Sciences, The University of British Columbia; Vancouver, Canada.

[2]Faculty of Medicine, Department Biochemistry & Molecular Biology, Michael Smith Laboratories, Life Sciences Institute, The University of British Columbia; Vancouver, Canada.

*Corresponding author. Email: dbromme@dentistery.ubc.ca

Abstract:

Proteases are classically viewed as degradative enzymes, yet they can also catalyze reverse proteolysis, a reaction known for decades but rarely examined in disease contexts. Here, we show that cysteine cathepsins efficiently form multi-generational fusion peptides, including chimeras between viral/bacterial and human proteins, revealing a hidden layer of peptidome complexity with direct relevance to autoimmunity. Importantly, we demonstrated that this activity is pH dependent, favored by citrullination, and may represent up to 4.5% of the protease activity. Moreover, we identified cathepsin S-generated type 1 diabetes superagonists, directly linking reverse proteolysis to a clinically established autoantigen. To capture these rare products in a cellular environment, we developed a click chemistry-based technology. These findings expand the protease functional paradigm and uncover a hidden source of disease-relevant antigens.

Introduction

Proteases are traditionally regarded as the cell's demolition crew, dismantling proteins into peptides. Yet the same catalytic machinery that drives peptide bond hydrolysis can also forge new bonds by stitching peptide fragments together. Despite being an inherent feature of protease chemistry (*1, 2*), this bidirectional reactivity has been largely overlooked in biology, where proteases are defined and studied primarily through the lens of degradation. This narrow view has left a blind spot concealing a potential universe of non-genomically encoded fusion peptides or even proteins with the potential to alter immune recognition. This oversight has begun to be corrected by recent findings showing that proteasomes can generate cis/transpeptides presented by MHC class I (*3-5*). More relevant to our study, evidence is emerging that MHC class II alleles implicated in type 1 diabetes (T1D), such as HLA-DQ8, present fusion peptides recognized by disease-specific CD4+ T cells. These peptides were either directly detected in the HLA-DQ8–bound immunopeptidome or synthesized to validate their HLA-DQ8 binding and T cell recognition (*6-8*). However, it remained unknown which proteases were responsible for their generation. Lysosomal cysteine cathepsins, being major contributors to antigen processing in antigen-presenting cells (APCs) (*9*), would be candidate proteases. Using *in vitro* assays, Reed et al., could demonstrate that Cathepsin L is able to produce highly antigenic chimeric epitopes for diabetogenic $CD4^+$ T cells from inactive fragments of a proteins (*10*). These preliminary results warrant a systematic investigation of the transpeptidase potential of lysosomal cysteine proteases. Although the underlying chemistry of reverse proteolysis is well established, there is limited knowledge of its substrate specificity, efficiency, and scope in biological systems. In particular, it remains unclear whether cysteine cathepsins can generate higher-order fusion peptides, fuse self- and non-self–derived fragments, or whether this activity is modulated by factors such as pH or post-translational modifications. Moreover, current methods for transpeptide discovery rely on classical proteomic approaches, using synthetic peptides and comparison of fragmentation ion patterns. However, there remains a lack of quantitative, high-throughput strategies for identifying transpeptides in live cells or *in vivo* settings.

In this research we have addressed several of these issues. We hypothesize that protease-catalyzed cycles of hydrolysis and re-fusion could generate large pools of cis/transpeptides that bridge pathogen- and host-derived fragments, thereby reshaping our understanding of immune surveillance at the host-host and host–microbe interface. This is especially relevant to infection-associated autoimmunity (*11, 12*), a field marked by strong epidemiological links between

viral/bacterial exposure and diseases like T1D, rheumatoid arthritis, and viral/vaccine-associated diseases such as thrombocytopenia (*13, 14*), for which mechanistic explanations are still incomplete.

We used three complementary approaches to address regulation, efficacy, and disease relevance of cathepsin-mediated cis/transpeptide formation. First, we conducted controlled *in vitro* enzyme assays, progressing from peptide–peptide, to peptide–protein, and then protein–protein interactions, to define the catalytic capacity of cysteine cathepsins to generate fusion peptides and elucidated the effect of pH and post-translational modifications on this activity. Second, we aimed to identify the yet unknown proteases which are responsible for generating the clinically observed MHC II-associated transpeptides, such as those reported in T1D (*6*). Third, because existing approaches lack quantitative, high-throughput strategies for identifying transpeptides in live cells or *in vivo*, we developed a targeted chemical enrichment strategy, Click-based Targeted Transpeptide Retrieval and Purification (CT-TRAP), and applied it in cell culture experiments to isolate fusion products.

Together, these approaches will show that lysosomal cysteine cathepsins can function iteratively, producing multi-generational cis/transpeptides, including host–infection-based fusion peptides, and that citrullination as well as autoimmune disease associated lysosomal alkalization can enhance this process. By expanding proteases' catalytic repertoire beyond degradation, our findings reframe cathepsins as dual-function molecular architects and establish reverse proteolysis as a previously under-valued generator of antigenic diversity with far-reaching implications for infection biology, autoimmunity, and diverse human diseases including ageing.

Results

Identification and quantification of multi-generational cis/transpeptides using a peptide substrate model assay. To investigate the balance between hydrolysis and transpeptidation, and the potential for higher-order products, we used human cathepsin S (hCatS) as it has been shown to be a key enzyme in MHCII associated antigen processing (*15-17*) and capable of peptide fusion (*18*).

We first investigated the role of free N-terminal peptides as nucleophiles in hCatS-catalyzed proteolysis at pH 6.5, which reflects the conditions in early endolysosomes of antigen presenting cells (*19*). Using a FRET-based substrate, Ac-AE(Edans)FRTTSQGGK-Dabcyl, we found that adding a non-fluorescent competitor peptide (NH_2-TTSQGGK-Biotin) increased the rate by 1.6-fold

(Fig. 1A). This effect was not due to enhanced proteolysis but by the competition with the NH_2-TTSQGGK-Dabcyl hydrolysis fragment in the reverse proteolysis reaction that would otherwise quench fluorescence. This was further supported by adding the quenched product itself (NH_2-TTSQGGK-Dabcyl), which suppressed the hydrolysis rate by >3-fold. These data demonstrate that the hydrolysis-transpeptidation equilibrium is sensitive to the concentration of free N-termini peptides.

Using LC-MS/MS, we next analyzed reaction products from hCatS incubation with the FRET substrate and a 10-fold molar excess of NH_2-TTSQGGK-Biotin. The peptides were designed to produce distinct signature ions for Edans, Dabcyl, and Biotin (Fig. S1), facilitating the identification of cis/trans-products. Our analysis revealed a complex mixture of first- to third-generation cis/trans-peptides (Fig. 1B, S2-S4). Kinetic analysis showed first-order trans-peptides peaked within 10 min before gradual re-hydrolysis (Fig. 1C). In contrast, second- and third-order products appeared after 30 min and accumulated over 5 h. Cis-peptides followed a similar trend. They were undetectable early on due to the absence of NH_2-TTSQGGK-Dabcyl. First-order cis-peptides (e.g., Ac-AE(Edans)FRTTTTSQGGK-Dabcyl) emerged after initial hydrolysis of the starting substrate, peaked at 30 min and declined after 2 h. Higher-order cis-peptides (e.g., TTTTTSQGGK-Biotin) formed after 1 h and continued to be generated up to 5 h (Fig. 1D). These data demonstrate that hCatS catalyzes a dynamic and iterative process involving a complex interplay of hydrolysis and transpeptidation.

Quantification using synthetic standards revealed that first-generation transpeptides reached up to 4.5% of the initial hydrolysis product (Ac-AE(Edans)FR-COOH) within 10 min (Fig. 1E). While these first-generation products were transient, a third-generation transpeptide accumulated over 5 h, reaching 0.1% of the total hydrolysis product and showed remarkable stability. The identity of these products was confirmed by matching retention times and fragmentation patterns with synthetic standards (Fig. 1F, S5-S6). Crucially, all observed hydrolysis and transpeptidase activities were abolished in the presence of the pan-cathepsin inhibitor E64, confirming that the reactions were indeed catalyzed by hCatS.

Effect of donor specificity and citrullination on cis/trans-peptidation by hCatS. To investigate if hCatS-mediated transpeptidation is a substrate-specific process, we used bovine serum albumin (BSA) as a protein acceptor and three distinct biotinylated peptides as nucleophilic donors, each

derived from a clinically relevant protein (hPF4, SARS-CoV-2 Spike protein, and islet amyloid polypeptide). While SDS-PAGE analysis showed no effect on the overall pattern of proteolytic fragments, we observed pronounced differences in the potency of each peptide to form biotinylated fusion products by hCatS (Fig. 2A, B). The hPF4-derived peptide (NH_2-TTSQGGK-Biotin) was at least 4 to 5 times more effective in forming transpeptides than the other two peptides (Fig. 2C). These results demonstrate that cathepsin-mediated transpeptidation is a highly donor peptide-specific process.

We next investigated the effect of citrullination, a post-translational modification on arginine residues linked to autoimmunity (*20*). Kinetic analysis showed that a citrullinated FRET-based substrate was hydrolyzed with a 2.7-fold higher catalytic efficiency by hCatS than its non-citrullinated counterpart (Fig. 2D). This enhanced cleavage also led to a significant increase in transpeptidation in the presence of the donor peptide, NH_2-TTSQGGK-Biotin, with most citrullinated hybrid peptides generated at levels 2-10 times higher than those from the non-citrullinated substrate (Fig. 2F-I, S9). Like the non-citrullinated version, the citrullinated substrate enabled the generation of multi-generational cis/trans-peptides (Fig. 2E, Fig. S7-S8). Kinetic analysis of selected products showed that the enhanced formation rates of citrullinated hybrid peptides resulted in increased levels of stable degradation products (Fig. 2J). As the citrullinated fragments primarily served as acceptor peptides, it can be concluded that cis/transpeptidation is also influenced by the acceptor portion of the fusion reaction. These results suggest that cathepsin-mediated transpeptidation is not merely a generic nucleophilic acyl-substitution reaction, but rather a substrate-specific process, with its preference potentially arising from steric and electronic interactions between the donor peptide's amino acid sequence and the enzyme's prime subsite.

***In vitro* identification of auto-antigenic transpeptides associated with T1D.** GQV**E**LGGGNAV**E**VLK, a fusion of peptides from proinsulin and islet amyloid polypeptide has been previously reported as a T1D autoantigen (*6*). Notably, the presence of Glu at positions 1 and 9 in this transpeptide are thought to contribute to its potential as a super-agonist due to its effective binding to the MHC II receptor expressed by the high-risk HLA-DQ8 allele (*21*). Although several reports have speculated on the involvement of cathepsins in this process (*22, 23*), specific cathepsins catalysing the formation of this transpeptide remain unknown. Thus, we sought to identify the specific cathepsin responsible for generating this transpeptide. We incubated human

proinsulin and a biotinylated donor peptide from the islet amyloid polypeptide (NH_2-NAVEGGK-Biotin) with hCatS to identify hybrid peptides. LC-MS/MS analysis confirmed the formation of various transpeptides (Fig. 3A), including ELGGGNAVEGGK-Biotin corresponding to the hybrid peptide identified by Tran et al. (*6*), with fragmentation ion patterns and retention times matching those of the synthetic control transpeptide (Fig. 3B and Fig. S11). The formation of this transpeptide was highly pH-dependent, with yields more than doubling at an endosomal pH of 6.5 when compared to a lysosomal pH of 4.5 (Fig. 3C). In addition to the known transpeptide, we discovered several novel fusion products (Fig. S12-S14) such as GIV**E**QCCTNAV**E**GGK-Biotin and ALYLNAVEGGK-Biotin (Fig. 3A). The generation of these diverse neopeptides, including a unique transpeptide with Gln at the P1 position (FVN***Q***HLCGNAV***E***GGK-Biotin, demonstrates hCatS's capacity to create a wide repertoire of potential autoantigens. Gln can be selectively deaminated by tissue transglutaminase, converting it to Glu, which may enhance its activity as a super-agonist (*24*). This transpeptide resembles those reported in other disease models like celiac disease (*24*) but has not been previously identified in epitopes/neoepitopes associated with T1D.

We next compared the transpeptidase activity of hCatS to other cysteine cathepsins, including the ubiquitously expressed hCatB and hCatL and the tissue-specific hCatV and hCatK. Human CatS produced the highest yield of the T1D transpeptide **ELGGG**NAVEGGK-Biotin, while the immune tissue-dominant cathepsins hCatS and hCatV showed higher formation rates for most identified transpeptides (Fig. 3D). hCatL demonstrated comparable potency only for a single transpeptide, **FVNQHLCG**NAVEGGK-Biotin. These findings demonstrate that while multiple cathepsins can catalyze transpeptidation, the repertoire of fusion products is highly enzyme-specific.

Given the epidemiological link between SARS-CoV-2 and T1D (*25, 26*), we investigated whether hCatS could generate hybrid antigens from human islet amyloid polypeptide donor peptide and SARS-CoV-2 Spike S1 and S2 subunits. We identified several biotinylated transpeptides, with the S2 subunit appearing to be a more efficient acceptor protein than S1 (Fig. 3E, Fig. S15-S16). The formation of these products was abolished by the cathepsin inhibitor E64. Crucially, we discovered the hybrid peptide **E**QDKNNAV**E**GGK-Biotin, which contains signature Glu residues at positions 1 and 9 known to create super-agonists for the HLA-DQ8 allele (*21*). This finding provides direct evidence of a transpeptidation event between a human and a viral protein, offering a plausible mechanistic link for virus-induced T1D.

Finally, we evaluated the binding affinity of four newly discovered transpeptides to HLA-DQ8, a key T1D risk allele. These newly identified transpeptides contained Glu residues at positions 1 and 9 and showed equal or higher binding affinity for HLA-DQ8 than the previously reported T1D autoantigen, GQV**E**LGGGNAV**E**VLK which also has appropriate Glu (*6*) (Fig. 3F). Strikingly, a 15-mer mimic of the insulin-derived transpeptide, **E**ALYLVCGNAV**E**VLK, displayed a 41-fold higher affinity to HLA-DQ8 than the known reference peptide. Similarly, the Spike-derived transpeptides, IAV**E**QDKNNAV**E**LK and YN**E**NGTITNAV**E**VLK exhibited a threefold higher or equal binding affinity even so the later peptide has Glu residues in 1 and 10 position (Fig. 3F). These results demonstrate that cathepsin-catalyzed transpeptidation can generate highly antigenic neo-peptides that bind strongly to a key disease-associated HLA allele.

Exploring cis/transpeptidation as a potential mechanism behind the induction of autoimmune diseases with unresolved etiology. To widen the scope of cathepsin-generated fusion peptides as triggers of autoimmune processes, we investigated whether hCatS could generate autoantigens in platelet factor 4 (hPF4)-associated autoimmune diseases, such as heparin-induced (HIT) and vaccine-induced immune thrombotic thrombocytopenia (VITT).

To test our hypothesis, we incubated full-length hPF4 with the hPF4-derived peptide NH_2-TTSQGGK-Biotin and hCatS at pH 6.5. We identified a diverse array of hybrid peptides via LC-MS/MS (Fig. 4A; Fig. S17-S19), including a third-generation product, **TTSTTTT**TTSQGGK-Biotin. The presence of all building block fragments in the reaction mixture confirmed that cathepsin-mediated transpeptidation can generate a diverse repertoire of complex antigens from a limited set of initial fragments (Fig. 4B). We further identified **SLEVIK**TTSQGGK-Biotin, a hybrid peptide containing part of a previously discovered VITT epitope (*27*). The formation of this peptide was completely abolished in the presence of the cathepsin inhibitor E64, confirming that the reaction was enzyme-catalyzed (Fig. 4C).

We next tested the formation of transpeptides between the hPF4-derived donor peptide and SARS-CoV-2 Spike subunits S1 and S2. Our analysis revealed biotinylated products, which were more prominent with the S2 subunit and increased at higher pH values (Fig. 4D, E; S20A), a finding consistent with our previous observations for T1D-associated transpeptides (Fig.3C). This pH-dependent activity also held true for full-length glycosylated Spike protein (Fig. S20B), demonstrating that this reaction is not limited to non-glycosylated substrates.

We also identified several biotinylated transpeptides from this reaction (Fig. 4F; Fig. S21-S24). Most notably, we discovered fusion peptides derived from the S1 receptor-binding domain (RBD) and the N-terminal domain (NTD) of the Spike protein, **FQPTNGVG**TTSQGGK-Biotin and **ALEPLVDLPIG**TTSQGGK-Biotin, respectively. Both are well-established antigenic regions (*28*). These findings suggest that transpeptides bridging highly antigenic peptide segments of hPF4, either within cis-peptides or with immunogenic motifs from viral Spike proteins in trans-peptides, may promote misrecognition of self and non-self, thereby contributing to PF4-associated thrombotic reactions.

Cis/Transpeptidation between full-length proteins: Exploring complex fusion peptides derived from SARS-CoV-2 Spike, proinsulin, and hPF4 proteins. To move further into physiologically relevant conditions, we next investigated whether hCatS could generate complex fusion peptides from full-length Spike subunits, proinsulin, and hPF4 at pH 6.5. Using LC-MS/MS and *de novo* peptide sequencing, we identified a diverse multitude of first- and second-generation cis- and transpeptides. Cis/transpeptides formed between insulin-insulin, spike-spike, and spike-insulin increased by up to 32,000-fold compared to control reactions (Fig. 5A). Analysis of cathepsin-inhibited reactions revealed a subset of cis- and transpeptides with significant fold changes and strong predicted HLA-DQ8 binding affinities (Fig. 5A).

Over 26 tandem mass spectra of novel transpeptides were discovered with a sequencing confidence exceeding 90%, and then manually verified. (Fig. 5B, Figs. S25–S34). We identified the Spike–Insulin transpeptide, **AENSVA**GSHLVE (Fig. 5C), a first-generation product that closely resembles a previously identified super-agonist with high-affinity binding for HLA-DQ8 (*29*). Furthermore, we identified the first examples of a second-generation cis-peptides, such as **LVCG**<u>*G*</u>FVNQHLC (Fig. 5D, Figs. S35 and S36), confirming the presence of all its precursor peptides in the reaction mixture.

To determine the potential of cathepsins to generate fusion peptides from full-length proteins, we compared the efficacy of hCatS, V, K, L, and B to form two 1$^{st.}$ cis-peptides (**GIVEQCCT**TPKT and **LVCG**GSHLVE). Similar to our findings for transpeptidation (Fig. 3D), hCatS and hCatV were the most effective proteases, while hCatL, K, and B produced only negligible amounts of these peptides (Fig. 5E).

Finally, we used full-length Spike and hPF4 proteins to identify fusion products relevant to thrombotic thrombocytopenia diseases like HIT and VITT. We identified 20 hPF4 cis-peptides and 60 hPF4-Spike transpeptides (Fig. 5F, S37-S40). Among these, a cis-derivative of a PF4-VITT epitope (**SLEVIK**LIATLK) (*27*) was of special interest, which was also identified in our previous experiment using the PF4-derived peptide donor (Fig. 4A, B).

Identification of Cis/Trans-peptides in cells using CT-TRAP. To investigate cis/transpeptidation in a biologically relevant context, we focused on antigen-presenting cells such as mouse RAW264.7 macrophages and splenic and thymic dendritic cells. For this purpose, we developed a new method, Click-based Targeted Transpeptide Retrieval and Purification (CT-TRAP) schematically shown in Figure 6A. The CT-TRAP probe was designed with a specific architecture: it included an N-terminal donor peptide containing the TTSQ motif, which acts as a competitive acceptor for transpeptidation. A Lys(N3) residue was incorporated as a click-chemistry handle, allowing for targeted enrichment of transpeptides and the generation of a robust marker ion via copper-catalyzed azide–alkyne cycloaddition (CuAAC). The CuAAC reaction produces a unique 4-PEG-biotin fragment ion (m/z 270.1276) that is used for identification because it does not overlap with any theoretical dipeptide. A fluorogenic 5FAM motif allowed monitoring of probe internalization, while the TAT-sequence (YGRKKRRQRRRK) enabled efficient lysosomal delivery. Using this probe (NH_2-TTSQGGK(N_3)-YGRKKRRQRRRK-5FAM), we confirmed up to 95% lysosomal uptake in RAW264.7 macrophages (Fig. 6A; Fig.S37).

Application of CT-TRAP combined with LC-MS/MS in RAW264.7 cells and *de novo* sequencing identified several peptides that incorporated the 4-PEG-Biotin tag (Fig. 6B; Fig. S38, S41), including the partially processed starting material, TTSQGGK(4PEG-Biotin)YGR. As a negative control, none of these peptides were detected in cell cultures lacking the tag peptide (Fig. 6B). The identification of fragments of the tag at both the *N*- and *C*-termini of the identified peptides indicated that the donor peptide underwent hydrolysis both before and after fusion (Fig. 6B, C). The complexity of these fusion and hydrolysis events was further highlighted by the identification of several fusion products containing only the Lysine-click-modified residue within their sequences (Fig. 6B), suggesting multi-generational processing. The tag peptides or fragments thereof fused with a variety of cellular proteins such as glyceraldehyde-3-phosphate dehydrogenase Fig. 6C). Importantly, the formation of these fusion products was abolished by E64d, as shown for an

individual peptide example (Fig. 6C). This inhibition by E64d clearly demonstrated that cysteine cathepsins were the responsible proteases generating this fusion product within the endolysosomal compartment.

Interestingly, the analysis of dendritic cells from mouse splenocytes revealed a transpeptide, **SHLVEALYLVC**FACPKFYTPTK, as a fusion product between mouse Insulin II and a peptide of bacterial origin (Fig. 6D; Fig. S42). This peptide was identified with a high *de novo* sequencing score of 96% (ALC%). Notably, a similar peptide was not observed in thymic DCs (Fig.6E). These results provide further evidence that cis/transpeptidation is a crucial, previously overlooked factor that may contribute to the generation of diverse neopeptides in immune cells.

Discussion

The immune system's ability to differentiate between "self" and "non-self" is critical for maintaining health. In autoimmune diseases, this distinction is lost, leading to the immune system attacking the body's own tissues (*30*). While theories like molecular mimicry and epitope spreading offer partial explanations for the initiation of these diseases (*31*), the exact mechanisms by which self-peptides become autoantigens remain a significant unknown in immunology.

Recent studies have explored protease-catalyzed transpeptidation, or "peptide splicing," as a potential source of neoantigens that could trigger autoimmunity. While the proteasome has been implicated in this process for MHC Class I presentation (*32-35*), evidence for the MHC Class II pathway has been less direct - either by identifying transpeptides in the immunopeptidome, by making synthetic derivatives of hybrid peptides capable of activating T-cells, or the *in vitro* generation of fusion products by a selected cathepsin (*7, 8, 10, 18*). The identification of MHC-class associated proteases in hybrid peptide formation as well as the characterization of factors influencing the formation of these peptides remained unsolved. Critics have argued against the physiological relevance of this process, suggesting that hydrolysis would overwhelmingly dominate over transpeptidation and that the existence of such products has not been conclusively proven *in vivo* (*36*). Our study addresses these criticism directly and revealed several important insights.

We used biotin-tagged peptides to unequivocally demonstrate that lysosomal cysteine cathepsins, specifically hCatS, can act as efficient iterative ligases. Our findings showed that hCatS can produce stable, multi-generational transpeptides. In a model system, transpeptidation accounted for up to 4.5% of the total reaction products, a yield sufficient to compete with hydrolysis and

relevant to the approximately 2% of transpeptides previously detected in HLA-DQ8 immunopeptidomes (*6*). Given that a single peptide-MHC complex can trigger a robust T-cell response, the levels of hybrid peptides we observed are physiologically significant and could easily initiate an immune response (*37, 38*).

It should be noted that the cis/transpeptidase activity yielded multi-generational fusion products due to subsequent hydrolysis and fusion events. Using short acceptor and donor peptides we identified up-to 3rd generation fusion products and interestingly, some of these products revealed increasing hydrolytic stability. The function of these peptides in a biological space remains completely unknown but could exhibit antigenic functions or lead to aggregation products as seen in aging and disease (*39*).

Cathepsin transpeptidase activity is highly pH-dependent, increasing significantly in less acidic conditions. This is particularly relevant as lysosomal dysfunction and associated alkalization are common in autoimmune diseases (*40*). This suggests that the generation of neoantigens through this mechanism may be enhanced in pathological states. As the optimal transpeptidase activity is at neutral pH, the extracellular generation of transproteins fragments by secreted cathepsins appears to be possible which then may directly act as antigens for B cells.

Of high relevance is also our discovery that post-translational modifications such as citrullination, can dramatically increase the efficiency of these transpeptidation reactions. This finding adds a new layer to the understanding of how citrullination contributes to autoimmune diseases like rheumatoid arthritis (*41*)., suggesting it not only alters the self/non-self boundary but also accelerates the creation of novel antigenic hybrid peptides.

Crucially, we demonstrated that hCatS can catalyze the formation of a known super-agonist peptide implicated in Type 1 Diabetes (*6*). This provides a direct mechanistic link between reverse proteolysis and a clinically relevant autoantigen. Furthermore, the ability of hCatS to generate a variety of other high-affinity fusion products, including those incorporating viral fragments such as from the Spike protein of SARS-CoV-2, highlights its broader role in shaping the antigenic repertoire and its potential involvement in infection-associated autoimmunity.

The cathepsin dependent specificities to generate individual fusion products are important as the distinct expression patterns of cathepsins in different immune tissues, such as the thymus versus the periphery (*17, 42, 43*), may lead to the generation of different sets of fusion products. This could allow self-reactive T cells that escaped negative selection to later encounter these unique hybrid peptides in the periphery and mount an autoimmune response. For example, it has been

shown that increased CatS expression in peripheral DCs contributes to Systemic lupus erythematosus (*44*).

Finally, using our CT-TRAP method allowed for the first time the discovery of highly diverse fusion peptides in antigen-presenting cells. This implies a multigenerational fusion and hydrolysis process by multiple lysosomal proteases present in dendritic cells and proves that cis/transpeptidation occurs in live cells.

In summary, our research establishes reverse proteolysis by lysosomal cysteine cathepsins as a fundamental mechanism for generating stable, higher-order cis/transpeptides. This expands the traditional view of antigen processing beyond simple degradation, providing a mechanistic basis for the formation of neoantigens from both self- and non-self proteins and aligns with the neoantigen hypothesis that disputes the notion that self-derived peptides that T cells develop central and peripheral tolerance for remain the same throughout life (*45*). By linking hCatS to the creation of a known disease-associated neoepitope and other high-affinity fusion products, our study provides a compelling framework for understanding the origins of autoimmunity. Furthermore, our findings suggest that targeting the formation of these neoantigens may offer new therapeutic strategies for treating autoimmune diseases. Future studies should focus on direct T-cell activation assays to confirm the immunogenicity of these hybrid peptides in a physiological context.

References

1. M. Bergmann, H. Fraenkel-Conrat, The role of specificity in the enzymatic synthesis of proteins - Syntheses with intracellular enzymes. *Journal of Biological Chemistry* **119**, 707-720 (1937).
2. K. Morihara, Using Proteases in Peptide-Synthesis. *Trends in Biotechnology* **5**, 164-170 (1987).
3. K. Hanada, J. W. Yewdell, J. C. Yang, Immune recognition of a human renal cancer antigen through post-translational protein splicing. *Nature* **427**, 252-256 (2004).
4. M. Mishto, A. Goede, K. T. Taube, C. Keller, K. Janek, P. Henklein, A. Niewienda, A. Kloss, S. Gohlke, B. Dahlmann, C. Enenkel, Driving forces of proteasome-catalyzed peptide splicing in yeast and humans. *Molecular & Cellular Proteomics* **11**, 1008–1023 (2012).
5. P. Faridi, A. Li, Z. Ramarathinam, L. Vivian, E. Reid, S. Song, D. G. G. Andrews, R. J. Purcell, C. Croft, T. W. Tscharke, J. Rossjohn, A. W. Purcell, A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Science Immunology* **3**, eaar3947 (2018).
6. M. T. Tran, J. E. Anderson, A. V. Sher, L. C. Huddleston, T. T. Smith, T. W. T. Petersen, S. R. Wallace, A. H. Balasubramanian, B. S. Jahnsen, E. R. Unanue, T. Michels, K. C. Herold,

D. A. Haskins, T cell receptor recognition of hybrid insulin peptides bound to HLA-DQ8. *Nature Communications* **12**, 5110 (2021).
7. Y. Wang, G. Sosinowski, S. A. Novikov, S. Crawford, R. R. White, D. M. Abad, J. J. Marré, D. A. Gerling, R. W. Paulsen, J. P. Krishnamurthy, J. W. Kappler, K. C. Herold, G. S. Eisenbarth, D. A. Haskins, C-terminal modification of the insulin B:11-23 peptide creates superagonists in mouse and human type 1 diabetes. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 162–167 (2018).
8. T. DeLong, R. Wiles, R. Baker, G. Bradley, J. Barbour, M. Reisdorph, K. Armstrong, S. Powell, R. Reisdorph, M. Kumar, K. W. Elso, J. DeNicola, C. Bottino, M. Trucco, G. S. Eisenbarth, J. W. Kappler, D. A. Haskins, Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science* **351**, 711–714 (2016).
9. J. A. Villadangos, R. J. Riese, C. Peters, H. A. Chapman, H. L. Ploegh, Proteases involved in MHC class II antigen presentation. *Immunological Reviews* **172**, 109–120 (1999).
10. B. Reed, J. Crawford, K. Hill, T. DeLong, J. K. Bradley, J. Marré, J. W. Kappler, D. A. Haskins, Lysosomal cathepsin creates chimeric epitopes for diabetogenic CD4 T cells via transpeptidation. *Journal of Experimental Medicine* **218**, e20192135 (2021).
11. P. R. Nielsen, T. W. Kragstrup, B. W. Deleuran, M. E. Benros, Infections as risk factor for autoimmune diseases—a nationwide study. *Journal of Autoimmunity* **74**, 176–181 (2016).
12. M. K. Smatti, H. Al-Sadeq, R. A. Ali, G. Pintus, R. Abou-Saleh, G. K. Nasrallah, Viruses and autoimmunity: A review on the potential interaction and molecular mechanisms. *Viruses* **11**, 762 (2019).
13. D. B. Cines, J. B. Bussel, SARS-CoV-2 vaccine-induced immune thrombotic thrombocytopenia. *The New England Journal of Medicine* **384**, 2254–2256 (2021).
14. C. Sharma, J. Bayry, High risk of autoimmune diseases after COVID-19. *Nature Reviews Rheumatology* **19**, 399–400 (2023).
15. K. Saegusa, T. Ishimaru, H. Yanagi, K. Arakaki, N. Ogawa, N. Saito, Y. Hayashi, Cathepsin S inhibitor prevents autoantigen presentation and autoimmunity. *The Journal of Clinical Investigation* **110**, 361–369 (2002
16. R. J. Riese, P. R. Wolf, D. Bromme, L. R. Natkin, J. A. Villadangos, H. L. Ploegh, H. A. Chapman, Essential role for cathepsin S in MHC class II–associated invariant chain processing and peptide loading. *Immunity* **4**, 357–366 (1996).
17. C. Stoeckle, A. Gouttefangeas, D. Hammerling, H. Arnold-Schild, H. L. Schild, H. Rammensee, S. Stevanović, Cathepsin S dominates autoantigen processing in human thymic dendritic cells. *Journal of Autoimmunity* **38**, 332–343 (2012).
18. Y. Wang, K. A. Scott, D. Brömme, Nucleophile labeling of cysteine and serine protease substrates. *The Journal of Biological Chemistry* **271**, 28399–28404 (1996).
19. E. Olson, M. Raghavan, Major histocompatibility complex class I assembly within endolysosomal pathways. *Current Opinion in Immunology* **84**, 102289 (2023).
20. J. Wang, J. Miao, P. Zhu, Insights into the complexities of citrullination: From immune regulation to autoimmune disease. *Autoimmunity Reviews* **24**, 103734 (2025).
21. K. N. Henderson, A. Tye-Din, H. A. Reid, J. A. Chen, D. L. Borg, P. A. Beissbarth, J. L. Tatham, M. J. Mannering, W. J. Purcell, J. Rossjohn, J. McCluskey, A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* **27**, 23–34 (2007).
22. Y. Wang, G. Sosinowski, S. A. Novikov, S. Crawford, R. R. White, D. M. Abad, J. J. Marré, R. W. Paulsen, J. P. Krishnamurthy, J. W. Kappler, D. A. Haskins, How C-terminal additions

to insulin B-chain fragments create superagonists for T cells in mouse and human type 1 diabetes. *Science Immunology* **4**, eaav7517 (2019).
23. S. I. Mannering, A. F. Rubin, R. Wang, P. Bhattacharjee, Identifying new hybrid insulin peptides (HIPs) in type 1 diabetes. *Frontiers in Immunology* **12**, 667870 (2021).
24. O. Molberg, S. N. McAdam, K. Körner, H. Quarsten, C. Kristiansen, L. Madsen, L. Fugger, H. Scott, E. Thorsby, L. M. Sollid, Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nature Medicine* **4**, 713–717 (1998).
25. B. I. Aydogan, U. Unluturk, M. Cesur, Type 1 diabetes mellitus following SARS-CoV-2 mRNA vaccination. *Endocrine***78**, 42–46 (2022).
26. A. S. Alsudais, A. A. Alaskar, F. A. Bin-Jumah, S. M. Alhumidi, Autoimmune diabetes mellitus after COVID-19 vaccination in adult population: A systematic review of case reports. *BMC Endocrine Disorders* **23**, 164 (2023).
27. E. Petito, R. Falcinelli, S. Paliani, S. Cesari, S. Vaudo, G. Sebastiano, A. Gresele, Association of human leucocyte antigen loci with vaccine-induced immune thrombotic thrombocytopenia: Potential role of the interaction between platelet factor 4–derived peptides and MHC-II. *British Journal of Haematology* **206**, 290–295 (2025).
28. A. Mittal, A. Khattri, V. Verma, Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathogens* **18**, e1010260 (2022).
29. A. Suri, J. J. Walters, M. L. Gross, E. R. Unanue, Natural peptides selected by diabetogenic DQ8 and murine I-A molecules show common sequence specificity. *The Journal of Clinical Investigation* **115**, 2268–2276 (2005).
30. P. A. Roche, K. Furuta, The ins and outs of MHC class II–mediated antigen processing and presentation. *Nature Reviews Immunology* **15**, 203–216 (2015).
31. R. Root-Bernstein, D. Fairweather, Unresolved issues in theories of autoimmune disease using myocarditis as a framework. *Journal of Theoretical Biology* **375**, 101–123 (2015).
32. J. Liepe, S. Mishto, P. Textoris-Taube, K. Janek, P. Henklein, A. Maldonado, G. Platzer, G. A. Mansurkhodzaev, S. J. Gomez-Perosanz, A. de la Rosa-Trevín, P. S. Ramos, D. R. P. Kloetzel, T. M. Demming, E. K. Fisette, S. Keller, The 20S proteasome splicing activity discovered by SpliceMet. *PLoS Computational Biology* **6**, e1000830 (2010).
33. J. Liepe, P. Marino, M. Sidney, J. Jeko, A. Bunting, K. Sette, A. S. Kloetzel, M. Mishto, A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science* **354**, 354–358 (2016).
34. M. Mishto, J. Liepe, Post-translational peptide splicing and T cell responses. *Trends in Immunology* **38**, 904–915 (2017).
35. J. Liepe, H. Ovaa, M. Mishto, Why do proteases mess up with antigen presentation by re-shuffling antigen sequences? *Current Opinion in Immunology* **52**, 81–86 (2018).
36. A. Admon, Are there indeed spliced peptides in the immunopeptidome? *Molecular & Cellular Proteomics* **20**, 100110 (2021).
37. J. Huang, P. Zarnitsyna, T. Liu, L. J. Edwards, S. Jiang, B. Evavold, C. Zhu, A single peptide–major histocompatibility complex ligand triggers digital cytokine secretion in CD4 T cells. *Immunity* **39**, 846–857 (2013).
38. R. Platzer, J. J. Stumpp, K. Reiser, M. Holzinger, C. Stutz, D. Schütz, P. Thiemann, P. B. Rosenthal, K. Dornmair, R. H. Herbst, E. S. Huseby, D. H. Busch, M. Minguet, V. Cerundolo, M. H. Meier-Schellersheim, W. W. A. Schamel, Monomeric agonist peptide/MHCII complexes activate T cells in an autonomous fashion. *EMBO Reports* **24**, e56649 (2023).

39. C. L. Klaips, G. G. Jayaraj, F. U. Hartl, Pathways of cellular proteostasis in aging and disease. *The Journal of Cell Biology* **217**, 51–63 (2018).
40. W. Zhang, Q. Wang, Y. Xu, L. Liu, H. Huang, X. Ma, Y. Zhao, Defective Lamtor5 leads to autoimmunity by deregulating v-ATPase and lysosomal acidification. *Advanced Science* **11**, e2400446 (2024)
41. A. M. Curran, M. K. Hensen, J. A. Weatheritt, K. C. Malinova, S. S. Makda, H. C. Cook, R. S. Goodyear, C. M. Sansom, J. Worthington, A. Skapenko, J. Marsal, K. A. Siminovitch, A. Barton, D. L. Scott, G. S. V. D. Geest, C. J. T. Filer, C. Holmdahl, D. L. Baeten, G. Schett, L. P. S. Radstake, J. M. W. Hazes, G. Steiner, E. Catrina, S. M. Anderton, I. E. Dumitriu, Citrullination modulates antigen processing and presentation by revealing cryptic epitopes in rheumatoid arthritis. *Nature Communications* **14**, 1061 (2023).
42. E. Tolosa, S. Li, J. Yasuda, R. Wienhold, T. Denzin, H. Lautwein, J. Driessen, C. Schnappauf, A. Kropshofer, H. A. Chapman, H. L. Ploegh, S. Melms, Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. *The Journal of Clinical Investigation* **112**, 517–526 (2003).
43. K. Mahiddine, C. Hassel, C. Murat, M. Girard, S. Guerder, Tissue-specific factors differentially regulate the expression of antigen-processing enzymes during dendritic cell ontogeny. *Frontiers in Immunology* **11**, 453 (2020).
44. S. J. Kim, J. Gregersen, H. G. Lee, E. J. Kim, Y. R. Choi, H. Y. Kim, J. Jang, J. W. Park, J. S. Choi, J. H. Lee, J. W. Yu, H. Y. Chung, K. H. Kim, Increased cathepsin S in Prdm1(−/−) dendritic cells alters the TFH cell repertoire and contributes to lupus. *Nature Immunology* **18**, 1016–1024 (2017).
45. T. Mustelin, F. Andrade, Autoimmunity: The neoantigen hypothesis. *Frontiers in Immunology* **15**, 1358419 (2024).

Figure legends

Fig.1: Cathepsin S-catalyzed multi-generational formation of cis/transpeptides using a model peptide substrate. (A) Time-course of hydrolysis of AC-AE(Ed)FRTTSQGGK(Dab) by hCatS in the presence or absence of NH_2-TTSQGG(Bio) or NH_2-TTSQGGK(Dab) at pH 6.5. (B) Schematic of hCatS-generated transpeptides at pH 6.5 identified by LC-MS/MS. (C,D) Time-course analysis of transpeptide generation (C) and cis-peptides (D) identified in (B). Peak areas were normalized to 100% at their maximal time points and plotted relative to that maximum. Comparisons are made within each peptide over time. (E) Percent formation of first-generation (Ac-AE(Ed)FRTTSQGGK-Bio) and third-generation (TTTTTSQGGK-Bio) transpeptides relative to first-generation hydrolysis product (Ac-AE(Ed)FR). (F) Comparison of retention time and fragment ion patterns between hCatS-generated Ac-AE(Ed)FRTTSQGGK(Bio) and its synthetic analogue (Bio: biotin; Dab: Dabcyl; Ed: Edans).

Fig. 2: Donor peptide specificity and citrullination effects on the cis/transpeptidase activity of human cathepsin S. (A) Schematic of the method used to detect biotinylated fusion peptides generated by cathepsins. (B) SDS-PAGE of Coomassie-positive hydrolysis and hybrid products and Avidin-FITC detected fusion products derived from BSA incubated with hCatS in the presence of three different biotinylated donor peptides at pH6.5 for 2 h at RT as well as in the presence or absence of the cathepsin inhibitor E64. (C) Quantification of FITC detectable biotinylated fusion products in the presence of three different donor peptides. * Biotinylated peptide was added after BSA hydrolysis and CatS inhibition to exclude non-specific donor peptide binding (BSA; bovine serum albumin). (D) Hydrolysis rates of Ac-AE(Ed)FRTTSQGGK(Dab) and Ac-AE(Ed)FCitTTSQGGK(Dab) by hCatS at pH 6.5 and their kinetic parameters in the accompanying table. (E) Proposed pathway for cis- and trans-peptide generation from Ac-AE(Ed)F*Cit*TTSQGGK(Dab) by hCatS. (F,G) Time-course formation of trans-peptides in independent reactions with citrullinated or non-citrullinated substrates. (H,I) Corresponding hydrolysis products from reactions in panels C and E. (J) Trans-peptide formation rates in panels F-I, calculated as peak area divided by time. (Bio: biotin; Dab: Dabcyl; Ed: Edans).

Fig. 3. Generation of putative TID super-agonists by cysteine cathepsins. (A) Selected biotinylated transpeptides identified in reaction mixture containing proinsulin and NH_2-NAVEGGK-

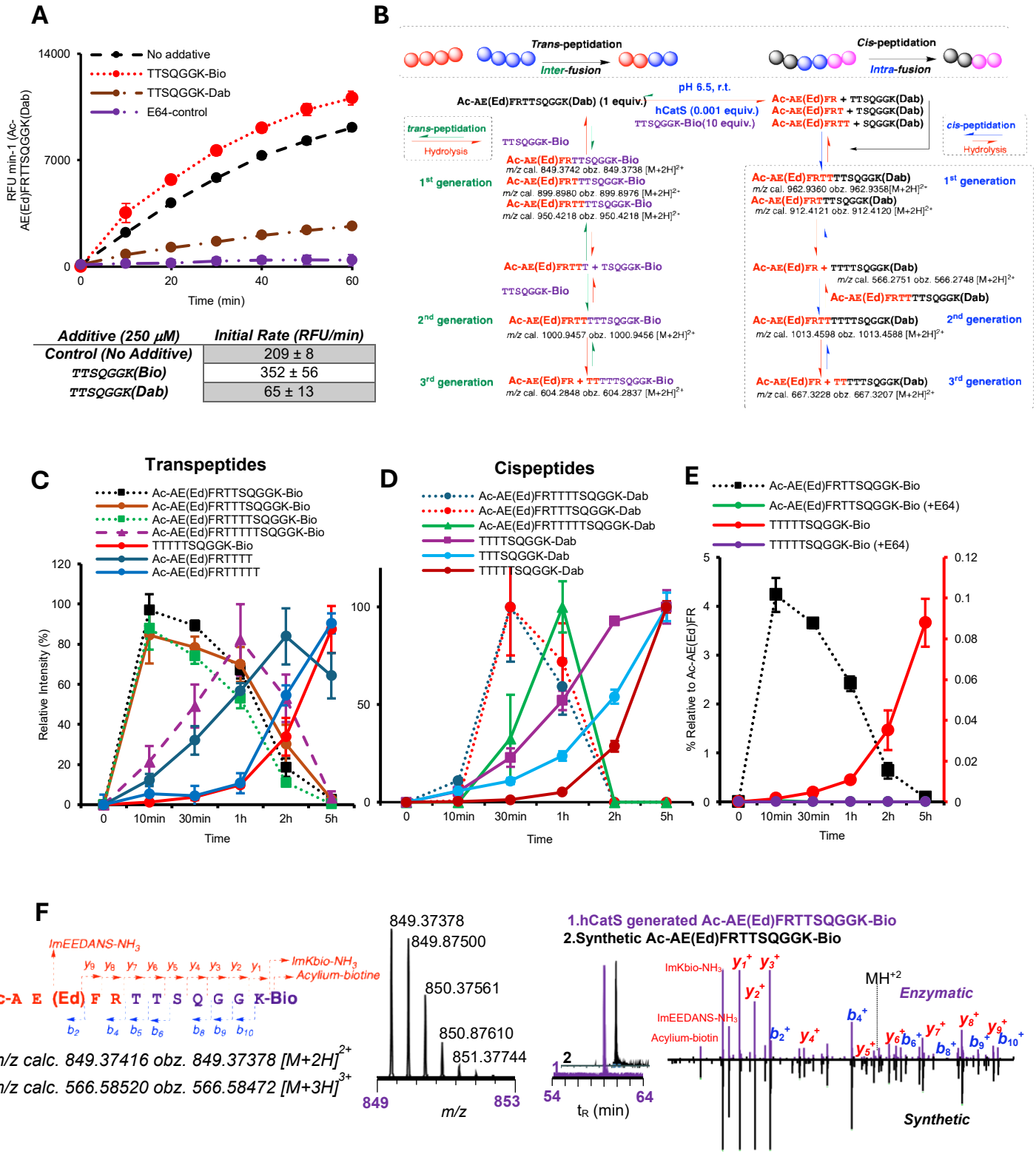
Biotin using LC-MS/MS and the location splicing site in then proinsulin sequence (in red). (B) Chromatogram comparing retention times of ELGGGNAVEGGK-Bio generated in reactions with hCatS in the absence or presence of E64. Fragment ion patterns of ELGGGNAVEGGK-Bio generated by hCatS and its synthetic analogue. (C) Peak areas of ELGGGNAVEGGK-Bio generated by hCatS at pH 4.5 and pH 6.5. (D) Biotinylated peptides identified by LC-MS/MS from reactions containing cathepsins at pH 6.5. (E) Avidin-FITC blot of SARS-CoV-2 Spike subunits (S1 and S2) incubated with hCatS and NH_2-NAVEGGK-Biotin at pH 6.5 for 2 h at RT. (*) indicates control experiments where NH_2-NAVEGGK-Biotin was added after a 2 h incubation of Spike subunits with active hCatS, followed by E64 inhibition. F) HLA-DQ8 binding assay (performed by ProImmune Ltd.) compares the binding capacity of 15-mer versions of discovered trans-peptides. (Bio: biotin).

Fig. 4. Generation of fusion peptides by hCatS in the presence of hPF4 or Spike protein subunits and a hPF4 derived antigenic biotinylated donor peptide. (A) Selected biotinylated cis/transpeptides identified in reaction mixture containing hPF4, NH_2-TTSQGGK-Bio and hCatS at pH 6.5 using LC-MS/MS and PF4 amino acid sequence with identified acceptor peptide regions in bold. (B) Ion chromatogram of SLEVIKTTSQGGK-Bio generated by hCatS in the presence or absence of E64. (C) Proposed mechanistic pathway for the formation of the identified third-generation transpeptide (TTSTTTTTTSQGGK-Bio). (D) Avidin-FITC blot of the reaction mixtures of Spike subunits (S1 and S2) incubated with hCatS and NH_2-TTSQGGK-Bio for 2 h at room temperature at different pH values. (E) Quantification of Avidin-FITC signals from (D). (F) Biotinylated transpeptides identified by LC-MS/MS from reaction mixtures containing Spike subunits (S1 and S2), NH2-TTSQGGK-Biotin, and hCatS at pH 6.5 for 2 h. (Bio: biotin).

Fig. 5. Fusion peptides generated by hcatS and other cysteine cathepsins from full-length proinsulin, hPF4, and S1/S2 subunit Spike proteins. (A) Scattered plot showing cis/transpeptides identified in reaction mixtures containing SARS-CoV-2 Spike protein, human proinsulin, and hCatS at pH 6.5 for 2 h at room temperature. (B) Selected 1$^{st.}$ generation cis/transpeptides generation by hCatS in the presence of Spike (S1, S2) protein and proinsulin. (C) Ion chromatograms of a selected transpeptide generated by hCatS in the presence or absence of E64. (D) Selected 1st. and 2nd generation cis-peptides derived from proinsulin by hCatS and their proposed pathway of formation. (E) Comparing the formation of two selected cis-peptides by different cysteine cathepsins. (F) 1$^{st.}$

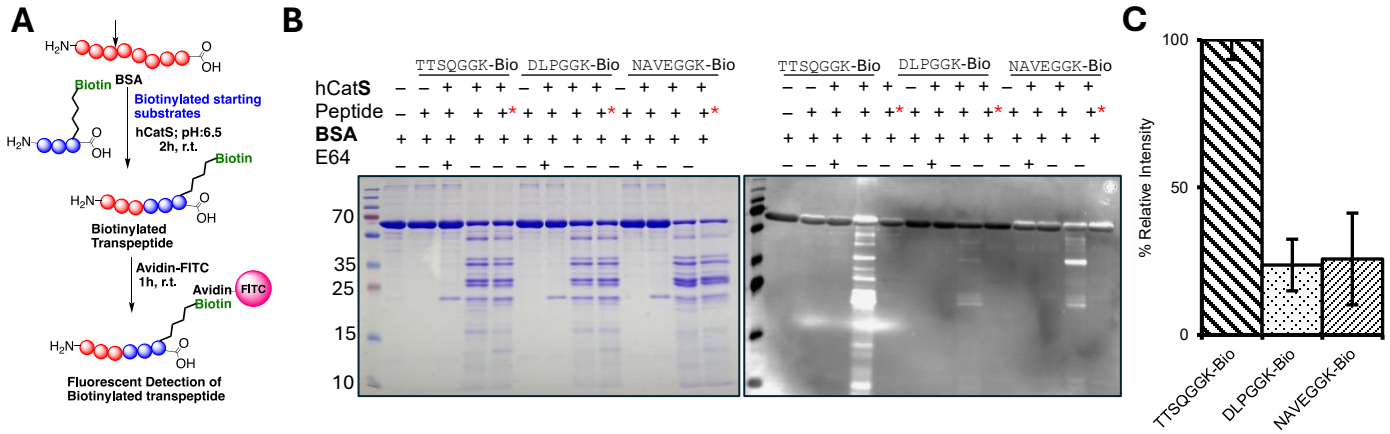
generation of cis/transpeptides identified from a reaction mixture containing Spike proteins, hPF4, and hCatS at pH 6.5 for 2 h at RT.

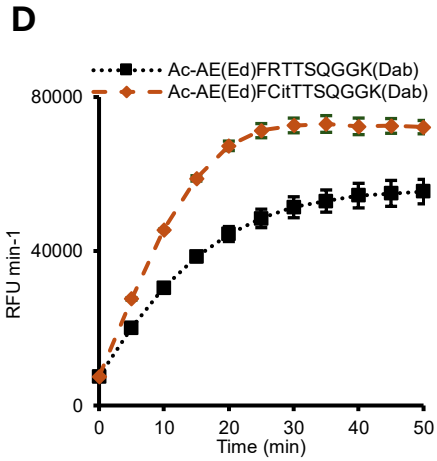
Fig. 6. Identification of fusion peptide in antigen-presenting cells (APCs). (A) Scheme of CT-TRAP method using a lysosomotropic TAT peptide (NH_2-TTSQGGK(N_3)-TAT-5FAM) and its cellular uptake to identify fusion products in cells. This method incorporated the inclusion of highly specific signature ions allowing the unequivocal identification of TAT-containing fusion products. (B) Selected TAT-peptide identified by LC-MS/MS containing the Lys-4PEG-Biotin signal at the *N*-, *C*-terminal as well as in the middle of the fusion peptides. (C) Ion chromatograms of a selected fusion peptide generated in RAW264.7 cells containing TAT-derived signature ions. This peptide was not identified when cells were treated with the cell-permeable cathepsin inhibitor, E64d. (D) Identified transpeptide between insulin and a bacterial protein found in splenic DCs from C57BL/6 mice. (E) Comparison of the presence of the insulin/bacterial transpeptide in splenic and thymic DCs (data from 2 mice).

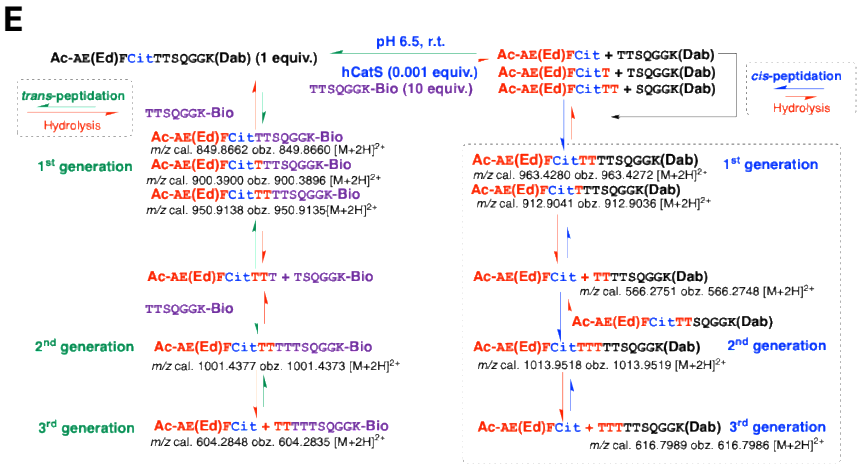


Additive (250 μM)	Initial Rate (RFU/min)
Control (No Additive)	209 ± 8
TTSQGGK(Bio)	352 ± 56
TTSQGGK(Dab)	65 ± 13

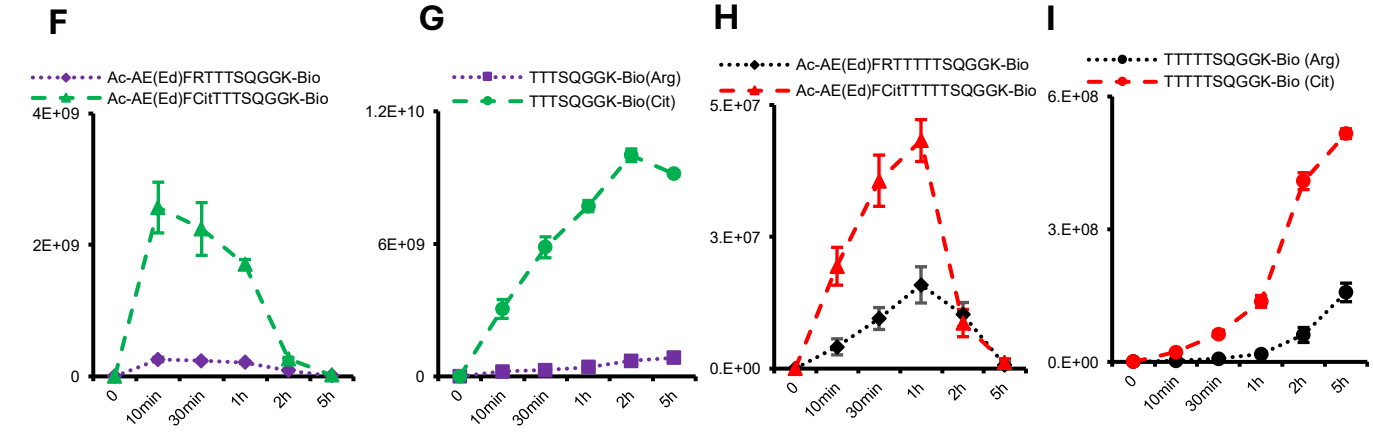
Figure 1







Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (μM^{-1} s^{-1})
Ac-AE(Ed)FRTTSQGGK(Dab)	13 ± 3	6 ± 0.8	0.43
Ac-AE(Ed)FCitTTSQGGK(Dab)	5.9 ± 0.9	7 ± 0.05	1.18



J

Parent peptide	Rate (PA/min)	Hydrolysis Product	Rate (PA/min)
Ac-AE(Ed)FRTTTSQGGK-Bio	2.3 ± 0.1 (x10^8)	TTTSQGGK-Bio	0.9 ± 0.2 (x10^8)
Ac-AE(Ed)FCitTTTSQGGK-Bio	22.4 ± 4 (x10^8)	TTTSQGGK-Bio	19.4 ± 1.5 (x10^8)
Ac-AE(Ed)FRTTTTTSQGGK-Bio	2.6 ± 0.5 (x10^5)	TTTTTSQGGK-Bio	5.2 ± 0.4 (x10^6)
Ac-AE(Ed)FCitTTTTTSQGGK-Bio	7.2 ± 0.6 (x10^5)	TTTTTSQGGK-Bio	17.2 ± 0.3 (x10^6)

Figure 2

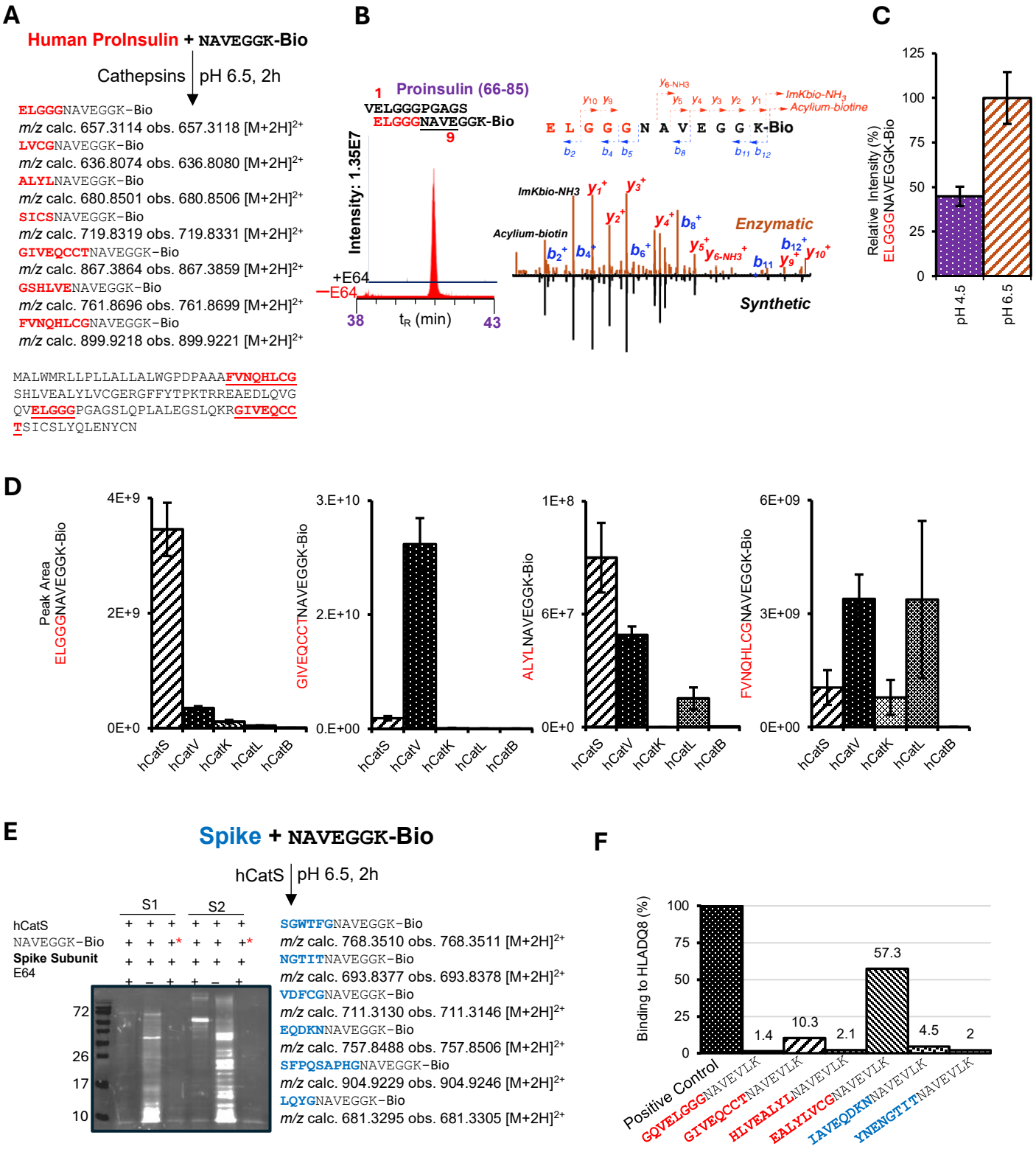

A
Human ProInsulin + NAVEGGK-Bio
Cathepsins
pH 6.5, 2h
ELGGGNAVEGGK-Bio
m/z calc. 657.3114 obs. 657.3118 [M+2H]2+
LVCGNAVEGGK-Bio
m/z calc. 636.8074 obs. 636.8080 [M+2H]2+
ALYLNAVEGGK-Bio
m/z calc. 680.8501 obs. 680.8506 [M+2H]2+
SICSNAVEGGK-Bio
m/z calc. 719.8319 obs. 719.8331 [M+2H]2+
GIVEQCCTNAVEGGK-Bio
m/z calc. 867.3864 obs. 867.3859 [M+2H]2+
GSHLVENAVEGGK-Bio
m/z calc. 761.8696 obs. 761.8699 [M+2H]2+
FVNQHLCGNAVEGGK-Bio
m/z calc. 899.9218 obs. 899.9221 [M+2H]2+
MALWMRLLPLLALLALWGPDPAAAFVNQHLCG
SHLVEALYLVCGERGFFYTPKTRREAEDLQVG
QVELGGGPGAGSLQPLALEGSLQKRGIVEQCC
TSICSLYQLENYCN
B
Proinsulin (66-85)
VELGGGPGAGS
ELGGGNAVEGGK-Bio
Intensity: 1.35E7
+E64
−E64
38
tR (min)
43
ImKbio-NH3
Acylium-biotin
Enzymatic
Synthetic
C
Relative Intensity (%)
ELGGGNAVEGGK-Bio
pH 4.5
pH 6.5
D
Peak Area
ELGGGNAVEGGK-Bio
GIVEQCCTNAVEGGK-Bio
ALYLNAVEGGK-Bio
FVNQHLCGNAVEGGK-Bio
hCatS
hCatV
hCatK
hCatL
hCatB
E
Spike + NAVEGGK-Bio
hCatS
pH 6.5, 2h
S1
S2
hCatS
NAVEGGK-Bio
Spike Subunit
E64
SGWTFGNAVEGGK-Bio
m/z calc. 768.3510 obs. 768.3511 [M+2H]2+
NGTITNAVEGGK-Bio
m/z calc. 693.8377 obs. 693.8378 [M+2H]2+
VDFCGNAVEGGK-Bio
m/z calc. 711.3130 obs. 711.3146 [M+2H]2+
EQDKNNAVEGGK-Bio
m/z calc. 757.8488 obs. 757.8506 [M+2H]2+
SFPQSAPHGNAVEGGK-Bio
m/z calc. 904.9229 obs. 904.9246 [M+2H]2+
LQYGNAVEGGK-Bio
m/z calc. 681.3295 obs. 681.3305 [M+2H]2+
F
Binding to HLADQ8 (%)
Positive Control
GQVELGGGNAVEVLK
GIVEQCCTNAVEVLK
HLVEALYLNAVEVLK
EALYLVCGNAVEVLK
LAVEQDKNNAVEVLK
YNENGTITNAVEVLK

Figure 3

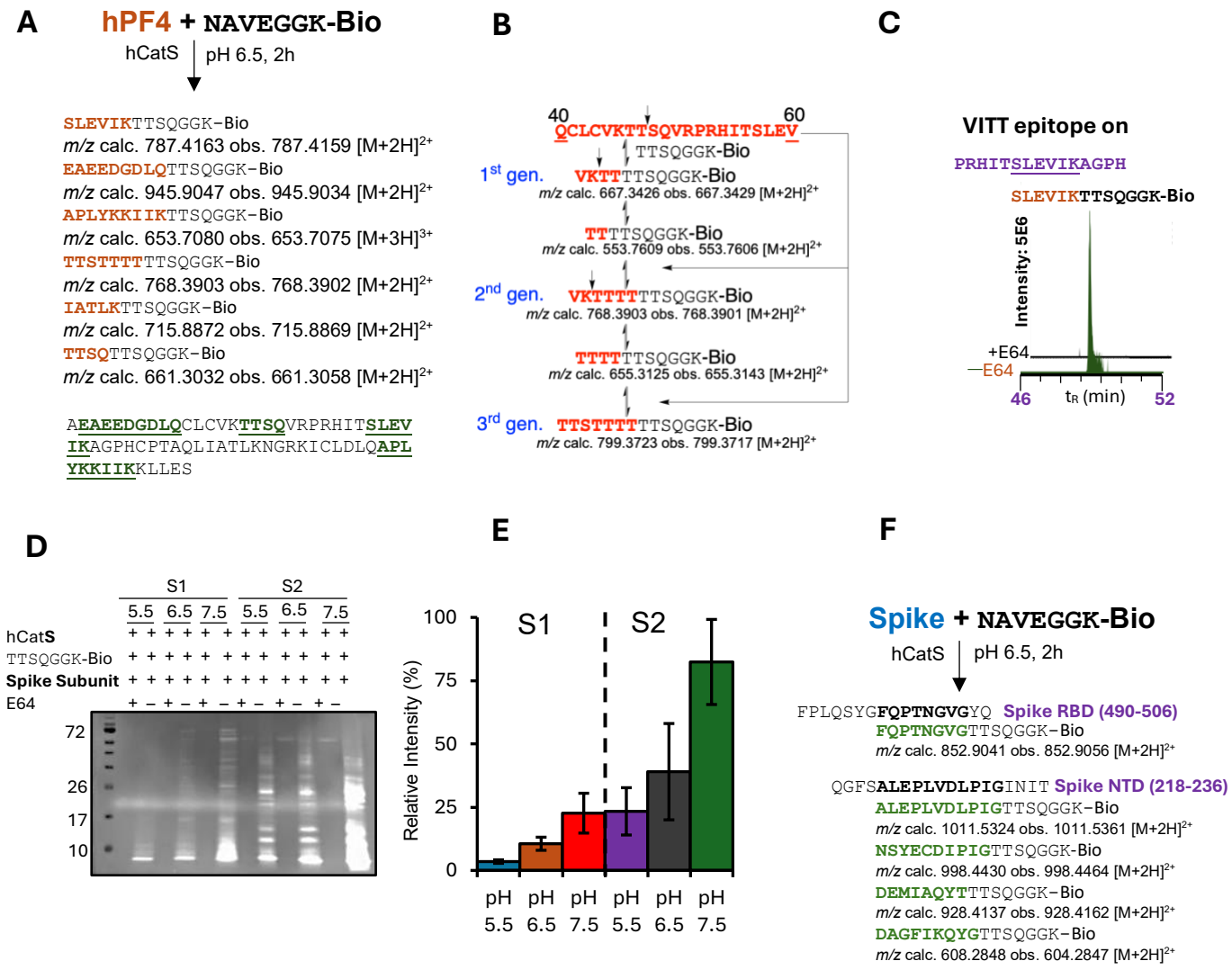

A
hPF4 + NAVEGGK-Bio
hCatS
pH 6.5, 2h
SLEVIKTTSQGGK-Bio
m/z calc. 787.4163 obs. 787.4159 [M+2H]2+
EAEEDGDLQTTSQGGK-Bio
m/z calc. 945.9047 obs. 945.9034 [M+2H]2+
APLYKKIIKTTSQGGK-Bio
m/z calc. 653.7080 obs. 653.7075 [M+3H]3+
TTSTTTTTTSQGGK-Bio
m/z calc. 768.3903 obs. 768.3902 [M+2H]2+
IATLKTTSQGGK-Bio
m/z calc. 715.8872 obs. 715.8869 [M+2H]2+
TTSQTTSQGGK-Bio
m/z calc. 661.3032 obs. 661.3058 [M+2H]2+
AEAEEDGDLQCLCVKTTSQVRPRHITSLEV
IKAGPHCPTAQLIATLKNGRKICLDLQAPL
YKKIIKKLLES
B
40
60
QCLCVKTTSQVRPRHITSLEV
TTSQGGK-Bio
1st gen.
VKTTTTSQGGK-Bio
m/z calc. 667.3426 obs. 667.3429 [M+2H]2+
TTTTSQGGK-Bio
m/z calc. 553.7609 obs. 553.7606 [M+2H]2+
2nd gen.
VKTTTTTTSQGGK-Bio
m/z calc. 768.3903 obs. 768.3901 [M+2H]2+
TTTTTTSQGGK-Bio
m/z calc. 655.3125 obs. 655.3143 [M+2H]2+
3rd gen.
TTSTTTTTTSQGGK-Bio
m/z calc. 799.3723 obs. 799.3717 [M+2H]2+
C
VITT epitope on
PRHITSLEVIKAGPH
SLEVIKTTSQGGK-Bio
Intensity: 5E6
+E64
E64
46
tR (min)
52
D
S1
S2
5.5 6.5 7.5 5.5 6.5 7.5
hCatS
TTSQGGK-Bio
Spike Subunit
E64
72
26
17
10
E
S1
S2
Relative Intensity (%)
100
75
50
25
0
pH 5.5
pH 6.5
pH 7.5
pH 5.5
pH 6.5
pH 7.5
F
Spike + NAVEGGK-Bio
hCatS
pH 6.5, 2h
FPLQSYGFQPTNGVGYQ Spike RBD (490-506)
FQPTNGVGTTSQGGK-Bio
m/z calc. 852.9041 obs. 852.9056 [M+2H]2+
QGFSALEPLVDLPIGINIT Spike NTD (218-236)
ALEPLVDLPIGTTSQGGK-Bio
m/z calc. 1011.5324 obs. 1011.5361 [M+2H]2+
NSYECDIPIGTTSQGGK-Bio
m/z calc. 998.4430 obs. 998.4464 [M+2H]2+
DEMIAQYTTTSQGGK-Bio
m/z calc. 928.4137 obs. 928.4162 [M+2H]2+
DAGFIKQYGTTSQGGK-Bio
m/z calc. 608.2848 obs. 604.2847 [M+2H]2+

Figure 4

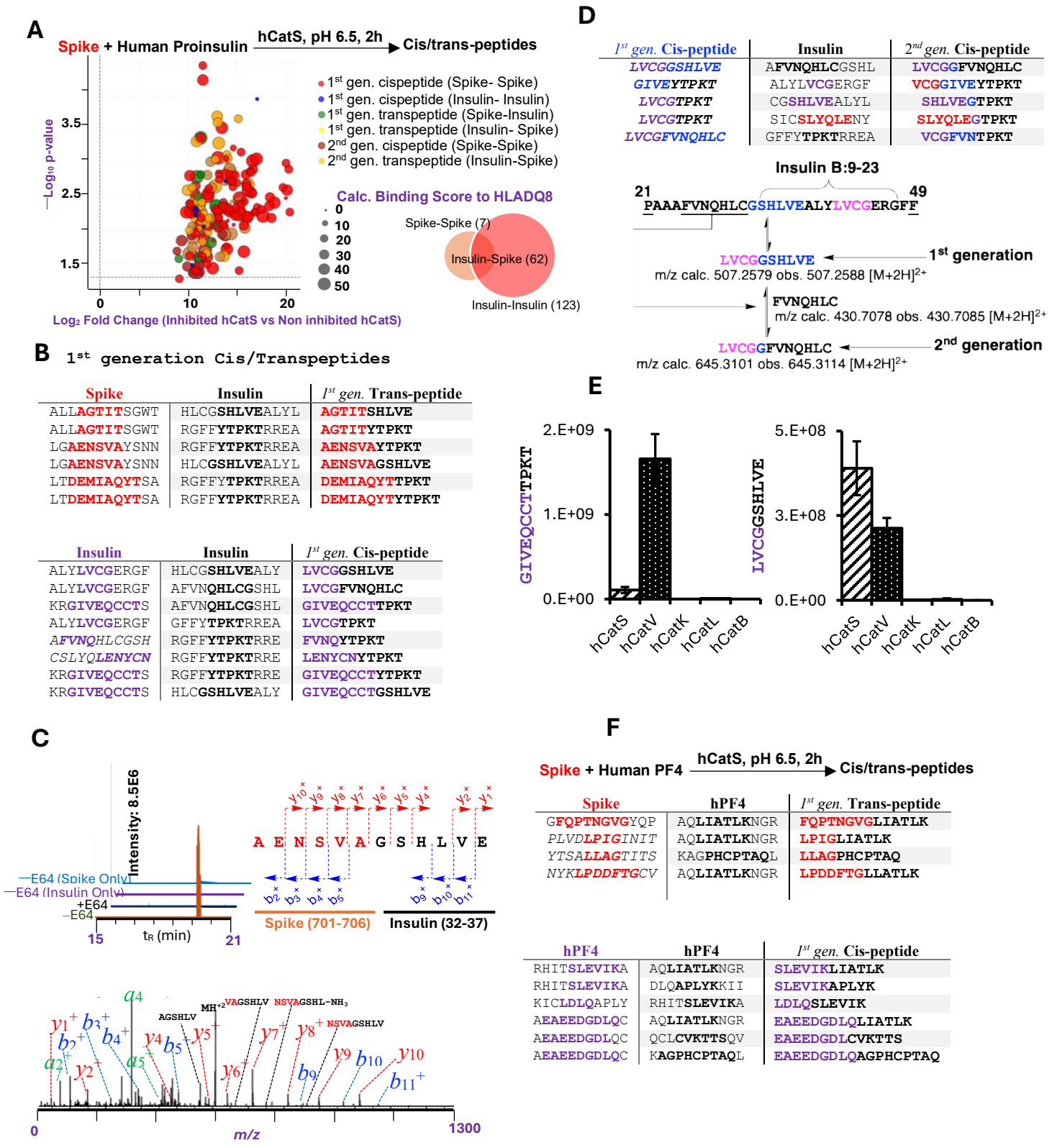

A
Spike + Human Proinsulin
hCatS, pH 6.5, 2h
Cis/trans-peptides
1st gen. cispeptide (Spike- Spike)
1st gen. cispeptide (Insulin- Insulin)
1st gen. transpeptide (Spike-Insulin)
1st gen. transpeptide (Insulin- Spike)
2nd gen. cispeptide (Spike-Spike)
2nd gen. transpeptide (Insulin-Spike)
Log10 p-value
Log2 Fold Change (Inhibited hCatS vs Non inhibited hCatS)
Calc. Binding Score to HLADQ8
Spike-Spike (7)
Insulin-Spike (62)
Insulin-Insulin (123)
B
1st generation Cis/Transpeptides
Spike | Insulin | 1st gen. Trans-peptide
ALLAGTITSGWT | HLCGSHLVEALYL | AGTITSHLVE
ALLAGTITSGWT | RGFFYTPKTRREA | AGTITYTPKT
LGAENSVAYSNN | RGFFYTPKTRREA | AENSVAYTPKT
LGAENSVAYSNN | HLCGSHLVEALYL | AENSVAGSHLVE
LTDEMIAQYTSA | RGFFYTPKTRREA | DEMIAQYTTPKT
LTDEMIAQYTSA | RGFFYTPKTRREA | DEMIAQYTYTPKT
Insulin | Insulin | 1st gen. Cis-peptide
ALYLVCGERGF | HLCGSHLVEALY | LVCGGSHLVE
ALYLVCGERGF | AFVNQHLCGSHL | LVCGFVNQHLC
KRGIVEQCCTS | AFVNQHLCGSHL | GIVEQCCTTPKT
ALYLVCGERGF | GFFYTPKTRREA | LVCGTPKT
AFVNQHLCGSH | RGFFYTPKTRRE | FVNQYTPKT
CSLYQLENYCN | RGFFYTPKTRRE | LENYCNYTPKT
KRGIVEQCCTS | RGFFYTPKTRRE | GIVEQCCTYTPKT
KRGIVEQCCTS | HLCGSHLVEALY | GIVEQCCTGSHLVE
C
Intensity: 8.5E6
E64 (Spike Only)
E64 (Insulin Only)
+E64
-E64
15
tR (min)
21
A E N S V A G S H L V E
Spike (701-706)
Insulin (32-37)
AGSHLV
MH+2
VAGSHLV
NSVAGSHL-NH3
NSVAGSHLV
0
m/z
1300
D
1st gen. Cis-peptide | Insulin | 2nd gen. Cis-peptide
LVCGGSHLVE | AFVNQHLCGSHL | LVCGGFVNQHLC
GIVEYTPKT | ALYLVCGERGF | VCGGIVEYTPKT
LVCGTPKT | CGSHLVEALYL | SHLVEGTPKT
LVCGTPKT | SICSLYQLENY | SLYQLEGTPKT
LVCGFVNQHLC | GFFYTPKTRREA | VCGFVNTPKT
Insulin B:9-23
21
49
PAAAFVNQHLCGSHLVEALYLVCGERGFF
LVCGGSHLVE
1st generation
m/z calc. 507.2579 obs. 507.2588 [M+2H]2+
FVNQHLC
m/z calc. 430.7078 obs. 430.7085 [M+2H]2+
LVCGGFVNQHLC
2nd generation
m/z calc. 645.3101 obs. 645.3114 [M+2H]2+
E
GIVEQCCTTPKT
LVCGGSHLVE
2.E+09
1.E+09
0.E+00
5.E+08
3.E+08
0.E+00
hCatS
hCatV
hCatK
hCatL
hCatB
F
Spike + Human PF4
hCatS, pH 6.5, 2h
Cis/trans-peptides
Spike | hPF4 | 1st gen. Trans-peptide
GFQPTNGVGYQP | AQLIATLKNGR | FQPTNGVGLIATLK
PLVDLPIGINIT | AQLIATLKNGR | LPIGLIATLK
YTSALLAGTITS | KAGPHCPTAQL | LLAGPHCPTAQ
NYKLPDDFTGCV | AQLIATLKNGR | LPDDFTGLLATLK
hPF4 | hPF4 | 1st gen. Cis-peptide
RHITSLEVIKA | AQLIATLKNGR | SLEVIKLIATLK
RHITSLEVIKA | DLQAPLYKKII | SLEVIKAPLYK
KICLDLQAPLY | RHITSLEVIKA | LDLQSLEVIK
AEAEEDGDLQC | AQLIATLKNGR | EAEEDGDLQLIATLK
AEAEEDGDLQC | QCLCVKTTSQV | EAEEDGDLCVKTTS
AEAEEDGDLQC | KAGPHCPTAQL | EAEEDGDLQAGPHCPTAQ

Figure 5

CT-TRAP: *Click-based Targeted Transpeptide Retrieval and Purification*

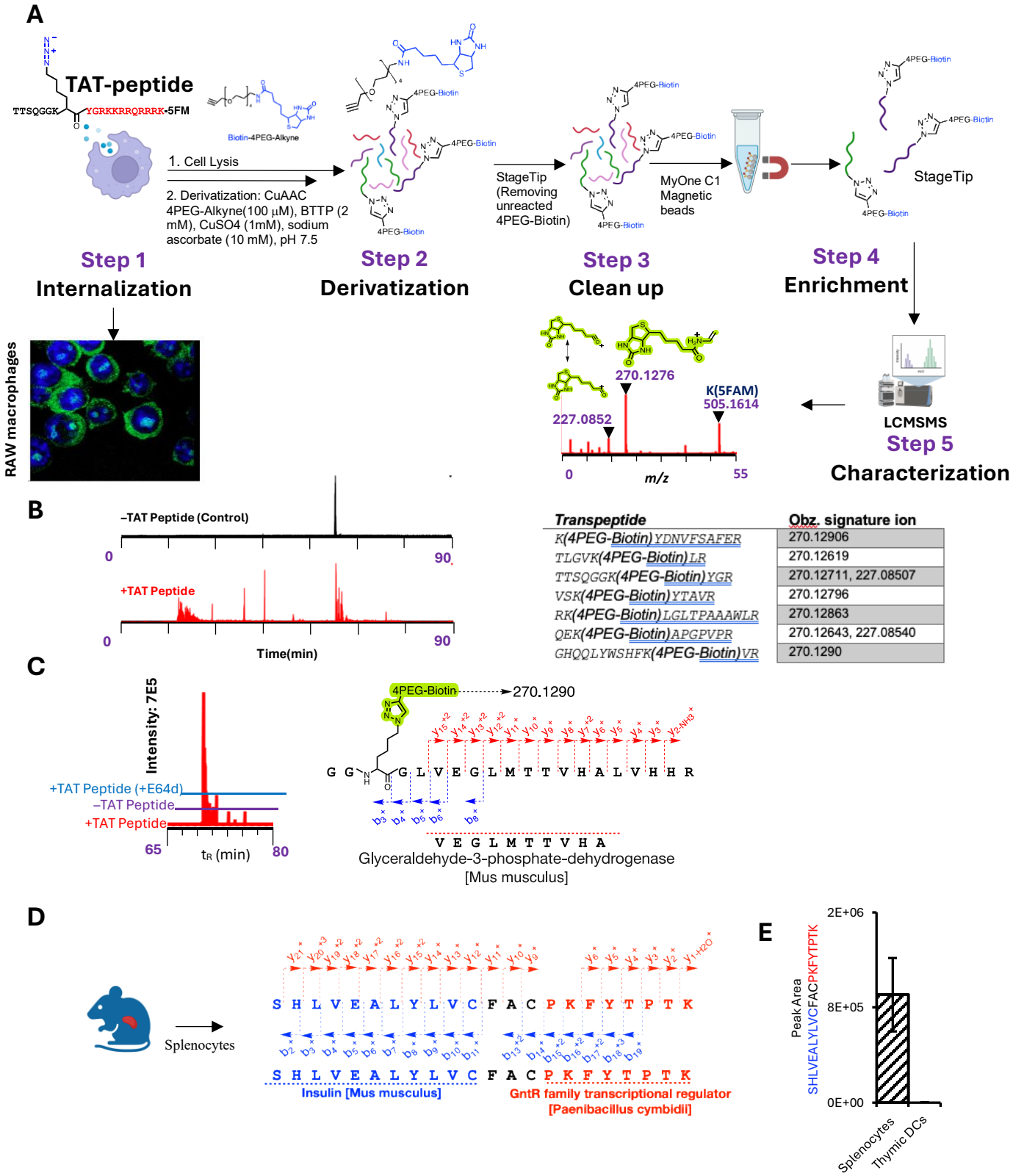


Figure 6