Reverse Proteolysis Uncovers a Hidden Dimension of the Peptidome

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 & 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 process known for decades but rarely examined in disease contexts. Here, we show that cysteine cathepsins efficiently form multi-generation cis/transpeptides, including chimeras between viral/bacterial, and human proteins, revealing a hidden layer of the peptidome complexity with direct relevance to autoimmunity. Importantly, we identified cathepsin S-generated type 1 diabetes super-agonists, directly linking reverse proteolysis to a clinically established autoantigen. Moreover, we demonstrated that this activity is pH dependent, favored by citrullination, and may represent up to 4.5% of total protease activity. To capture these 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 conventionally regarded as degradative enzymes that dismantle proteins into smaller fragments, yet the same catalytic machinery can also create new peptide bonds by joining peptide fragments. Although widely used in peptide chemistry (*1*), this reactivity has been largely overlooked in peptidomics, leaving protease-catalyzed fusion peptides (cis/transpeptides) underappreciated.

Recent studies have shown that proteasomes can generate cis/transpeptides presented by major histocompatibility complex (MHC) MHC class I molecules (*2-4*). Similarly, MHC class II molecules implicated in type 1 diabetes (T1D), such as human leukocyte antigen (HLA) HLA-DQ8, have been shown to present cis/transpeptides that are specifically recognized by disease-associated $CD4^+$ T cells (*5*). Despite these advances, biochemical validation remains limited, leading to skepticism that the reported cis/transpeptides may instead represent mass-spectrometric artefacts, mRNA splicing products, or genomic mutations. (*6*). Reverse proteolysis continues to be poorly defined regarding enzyme-level characteristics, including substrate specificity, catalytic efficiency, higher-order cis/transpeptide generation, and modulation by post-translational modifications.

As cysteine cathepsins are key proteases in MHC class II-linked antigen processing and presentation (*7*), we hypothesize that these enzymes are prime catalysts of cis/transpeptides, given their abundance, broad specificity, autoimmune relevance, and activity across a relatively wide pH range. We propose that cathepsin-mediated cycles of hydrolysis and re-fusion could generate diverse repertoires 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 (*8, 9*), a field marked by strong epidemiological links between viral/bacterial exposure and diseases like T1D, rheumatoid arthritis, and viral/vaccine-associated thrombocytopenia (*10, 11*), for which mechanistic explanations are still incomplete. Moreover, existing proteomic approaches lack quantitative, high-throughput strategies for identifying cis/transpeptides in live cells or *in vivo*.

To address these challenges, we employed three complementary approaches. First, we conducted controlled *in vitro* enzyme assays, progressing from small synthetic tagged peptide–peptide, to peptide–protein, and then protein–protein interactions, to define the

catalytic capacity of cysteine cathepsins to generate cis/transpeptides. We also elucidated the effect of pH and post-translational modifications on this activity. Second, we aimed to identify the unknown proteases responsible for generating the clinically observed MHC class II-associated transpeptides, such as those reported in T1D patients, thereby making the experiments and subsequent peptide discoveries more clinically relevant and mechanistically informative (*5*). Third, to address gaps in current methodological approaches for transpeptide discovery, we developed a targeted chemical enrichment strategy, Click-based Targeted Transpeptide Retrieval and Purification (CT-TRAP), and applied it to a murine macrophage cell line to isolate cis/transpeptides.

Together, these approaches demonstrate that lysosomal cysteine cathepsins can function iteratively, producing multi-generation cis/transpeptides, including host-viral products, and that citrullination as well as autoimmune disease-associated lysosomal pH alterations can enhance this process. By expanding the catalytic repertoire of proteases 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 and aging.

Results

Identification and quantification of multi-generational cis/transpeptides using a tagged synthetic peptide substrate assay. To investigate the balance between hydrolysis and cis/transpeptidation, and the potential for higher-order products, we used human cathepsin S (hCatS) because it has been shown to be a key enzyme in MHC class II-associated antigen processing (*12*-*14*) and is capable of cis/transpeptidation (*15*).

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 (APCs) (*16*). Using a Förster resonance energy transfer (FRET)-based substrate, Ac-AE(Edans)FRTTSQGGK-Dabcyl, we found that adding a non-fluorescent competitor peptide (NH_2-TTSQGGK-Biotin) increased the rate of reaction by 1.6-fold (Fig. 1A). This effect was not due to enhanced proteolysis but rather to 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 results demonstrate that the hydrolysis-cis/transpeptidation equilibrium is dependent on the concentration of free *N*-termini peptides.

Using LC-MS/MS, we then analyzed reaction products from hCatS incubated with a 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), thereby enhancing confidence in the identification of cis/transpeptides. Our analysis revealed a mixture of first- to third-generation cis/transpeptides (Fig. 1B, S2-S4). Kinetic analysis showed first-order transpeptides peaked within 10 min, before undergoing gradual re-hydrolysis (Fig. 1C). In contrast, second- and third-order products appeared after 30 min and accumulated over 5 h. Cispeptides followed a similar trend. They were undetectable early on due to the absence of NH_2-TTSQGGK-Dabcyl. First-order cispeptides (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 cispeptides (e.g., TTTTTSQGGK-Biotin) formed after 1 h and continued to be generated up to 5 h (Fig. 1D). These results demonstrate that hCatS catalyzes a dynamic and iterative process involving a complex interplay of hydrolysis and transpeptidation.

Since the immunological relevance of transpeptides depends not only on their identity but also on their abundance, we performed quantitative analysis of selected transpeptides using synthetic standards. This revealed that a selected first-generation transpeptide (AE(Edans)FR-TTSQGGK(Biotin)) reached up to ~4.5% of one of the initial hydrolysis products (Ac-AE(Edans)FR-COOH) within 10 min (Fig. 1E). While these first-generation products were transient, a third-generation transpeptide (TTTTTSQGGK-Biotin) accumulated over 5 h, reaching 0.1% of the total hydrolysis product and was notably resistant to further hydrolysis. 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 E-64, confirming that the reactions were indeed catalyzed by hCatS.

Effect of nucleophilic peptide specificity and citrullination on cis/transpeptidation by hCatS. To investigate whether hCatS-mediated transpeptidation is a substrate-specific process, we used bovine serum albumin (BSA) as an acceptor substrate and three distinct biotinylated nucleophilic donor peptides, each derived from a clinically relevant protein (human platelet factor 4 (hPF4)), SARS-CoV-2 spike (S) protein, or islet amyloid polypeptide). While SDS-PAGE analysis showed no effect of biotinylated peptides on the overall pattern of proteolytic fragments, we observed distinct differences in the potency of each synthetic biotinylated peptide in forming transpeptides by hCatS (Fig. 2A, B). The hPF4-derived peptide (NH_2-TTSQGGK-Biotin) was at least 4 to 5 times more efficient in forming transpeptides than the other two peptides (Fig. 2C). These results demonstrate that cathepsin-mediated transpeptidation is a strongly nucleophilic-peptide-specific process.

Because citrullination of Arg residues is a post-translational modification strongly associated with autoimmunity and known to alter protease recognition, we then investigated its effect on hCatS-mediated hydrolysis and transpeptidation (*17*). 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 analogue (Fig. 2D). This enhanced cleavage also led to a substancial increase in transpeptidation in the presence of the nucleophilic peptide, NH_2-TTSQGGK-Biotin, with most citrullinated cis/transpeptides generated at levels 2- to 10-fold 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-generation cis/transpeptides (Fig. 2E, Fig. S7-S8). Kinetic analysis of selected products showed that the accelerated formation of citrullinated cis/transpeptides led to increased abundance of stable cis/transpeptides (Fig. 2J). 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 arising from steric and electronic interactions between the nucleophilic peptide's amino acid sequence and the enzyme's prime subsite. Furthermore, it also depends on the susceptibility of the original substrate for hydrolysis.

***In vitro* identification of auto-antigenic transpeptides associated with T1D.** To investigate the transpeptidation catalytic activity of cysteine cathepsins in the context of autoimmune diseases, we focused on a previously established transpeptide implicated in the development of T1D. This transpeptide **GQVELGGG**NAVEVLK, resulted from the fusion of a proinsulin-derived peptide (**GQVELGGG**) and a peptide derived from the islet amyloid polypeptide (NAVEVLK) (*5*). It has been proposed that these antigenic transpeptides evade negative selection in the thymus and trigger an autoimmune response. Notably, the presence of Glu at positions 1 and 9 in this transpeptide is thought to contribute to its potential as a super-agonist due to its effective binding to the MHC class II receptor expressed by the high-risk HLA-DQ8 molecules (*18*). Although several reports have speculated on the involvement of cathepsins in this process (*19, 20*), and there is some experimental proof that cathepsin L is capable of forming related cis/transpeptides (*21*), the wider-range capability of cathepsins catalyzing the formation of this and, yet unknown related neoantigenic transpeptides, along with how it is affected by the endolysosomal pH environment remains poorly understood. Thus, we sought to identify the specific cathepsin responsible for generating this transpeptide and related transpeptides. We reacted human proinsulin and a biotinylated nucleophilic peptide from the islet amyloid polypeptide (NH_2-NAVEGGK-Biotin) with hCatS to identify cis/transpeptide. LC-MS/MS analysis confirmed the formation of various transpeptides (Fig. 3A), including **ELGGG**NAVEGGK-Biotin, which corresponds to the cis/transpeptide previously reported (*5*), with fragmentation ion patterns and retention times matching those of the synthetic analogue (Fig. 3B and Fig. S11). The formation of this transpeptide was pH-dependent, with yields more than doubling at endosomal pH 6.5 compared to lysosomal pH 4.5 (Fig. 3C). In addition to the known transpeptide, we discovered several novel transpeptides (Fig. S12-S14) such as **GIV*E*QCCT**NAV*E*GGK-Biotin and **ALYL**NAVEGGK-Biotin (Fig. 3A). The generation of these diverse neopeptides, including a unique transpeptide with Gln at the P1 position (**FVN*Q*HLCG**NAV*E*GGK-Biotin), demonstrates hCatSs 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 (*22*). This transpeptide resembles those reported in other disease models, like celiac disease (*22*), but has not been previously identified as an epitope/neoepitope associated with T1D.

To determine whether cis/transpeptide generation is a general property of cysteine cathepsins or a unique feature of hCatS, we compared its transpeptidase activity with that of other family members, including the ubiquitously expressed hCatB and hCatL and the tissue-specific hCatV and hCatK. This comparison is crucial, because hCatS is a prime candidate for driving autoimmune neoepitope formation in APCs, while other cathepsins could catalyze cis/transpeptides in distinct tissue environments or pathological settings, thereby broadening the scope of cis/transpeptide biology beyond autoimmunity. Human CatS produced the highest yield of the T1D transpeptide **ELGGG**NAVEGGK-Biotin. It also appears that immune tissue-dominant cathepsins hCatS and hCatV showed higher formation rates for most identified transpeptides in contrast to the ubiquitously expressed hCatL and hCatB (Fig. 3D). Human CatL demonstrated comparable potency for only one of the detectable transpeptides, **FVNQHLCG**NAVEGGK-Biotin. These findings demonstrate that while multiple cathepsins can catalyze transpeptidation, the repertoire of cis/transpeptides is distinct and enzyme-specific.

Given the epidemiological link between SARS-CoV-2 infection and vaccination and T1D (*23-25*), we investigated whether hCatS could generate cis/transpeptides from human islet amyloid polypeptide-derived biotinylated nucleophilic peptide and SARS-CoV-2 S1 and S2 subunits. We identified several biotinylated transpeptides, with the S2 subunit appearing to be a more efficient acceptor substrate than S1 (Fig. 3E, Fig. S15-S16). The formation of these products was stopped by the cathepsin inhibitor E-64. Crucially, we discovered the cis/transpeptide ***E*QDKN**NAV*E*GGK-Biotin, which contains signature Glu residues at positions 1 and 9 known to create super-agonists for HLA-DQ8 molecules (*18*). 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 five newly discovered transpeptides for HLA-DQ8, a key T1D risk factor. Four of them contained Glu residues at positions 1 and 9 and showed equal or higher binding affinity for HLA-DQ8 than the previously reported T1D autoantigen (*5*) (Fig. 3F). Notably, a 15-mer mimic of an insulin-derived transpeptide, **EAL*Y*LVCG**NAV*E*VLK, displayed a 41-fold higher affinity for HLA-DQ8 than the known reference peptide. Similarly, two spike-derived transpeptides, **IAV*E*QDKN**NAV*E*LK and **YNE*N*GTIT**NAV*E*VLK, exhibited a threefold higher or equal binding affinity,respectively,

although the latter has Glu residues at positions 1 and 10 (Fig. 3F). These results demonstrate that cathepsin-catalyzed transpeptidation can generate highly antigenic neopeptides that bind strongly to a key disease-associated HLA molecule.

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

To investigate this, we selected a traceable nucleophilic peptide (NH_2-**TTSQ**GGK–Biotin) sequence located within a major PF4 epitope known to trigger thrombotic autoimmunity (HIT and VITT; **LQCLCVK↓TTSQVRPRH**) (*26, 27*), and incubated it with full-length hPF4 in the presence of hCatS at pH 6.5. We identified a diverse array of cis/transpeptides using LC-MS/MS (Fig. 4A; Fig. S17-S19), including a third-generation cispeptide, **TTSTTTT**TTSQGGK-Biotin. The presence of all intermediate fragments in the reaction mixture confirmed that cathepsin-mediated transpeptidation can generate a diverse repertoire of potential antigens from a limited set of initial fragments (Fig. 4B). We also identified **SLEVIK**TTSQGGK-Biotin, a cispeptide containing part of a previously discovered VITT epitope (*28*). The formation of this peptide was completely abolished in the presence of the cathepsin inhibitor E-64, confirming that the reaction was hCatS-catalyzed (Fig. 4C).

Next, we tested the formation of transpeptides between the hPF4-derived nucleophilic peptide and SARS-CoV-2 spike subunits S1 and S2. Our results revealed the formation of higher molecular weight biotinylated products which were more prominent in the presence of the S2 subunit and increased for both subunits with pH (Fig. 4D, E; S20A), a finding consistent with our previous observations for T1D-associated transpeptides (Fig. 3C). Most notably, we discovered cis/transpeptides 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 (*29*). This hCatS-orchestrated transpeptidase activity was also observed

with the full-length glycosylated spike protein (Fig. S20B), demonstrating that this activity is not confined to *E.coli*-expressed, non-glycosylated spike subunits but extends to physiologically relevant viral proteins.

These findings suggest that the formation of such transpeptides, bridging highly antigenic regions of hPF4 within cispeptides or hPF4 and a viral protein in transpeptides, may have the potential to contribute to immune misrecognition of self and foreign antigens, thereby triggering the PF4-associated thrombotic reactions seen during the pandemic and the vaccination campaign.

Cis/Transpeptidation between full-length proteins: Exploring complex cis/transpeptides 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 cis/transpeptides from full-length Spike subunits, proinsulin, and hPF4. Using LC-MS/MS and *de novo* peptide sequencing, we identified a diverse array of first- and second-generation cis/transpeptides. The analysis identified several high-confidence cis/transpeptides formed from insulin and spike proteins, reproducibly detected in all three experimental replicates but absent in controls (Fig. 5A). The highest-ranked peptide as a HLA-DQ8 binder with a rank of 9 was identified for the spike-insulin transpeptide **DEMIAQYT**SHLVEAL. The potential immunogenicity of these peptides was assessed using a percentile-based ranking system from the Immune Epitope Database (IEDB) (*30*).

Over 26 tandem mass spectra of novel transpeptides were discovered with a sequencing confidence exceeding 90%, and were manually verified. (Fig. 5A, Figs. S25–S34). We identified the spike–insulin transpeptide, **DEMIAQYT**FVNQHLC (Fig. 5B; Fig. S38), a first-generation product with a high *de novo* sequencing score of ALC 96.6% (ALC%) that closely resembles a previously identified super-agonist with high-affinity binding for HLA-DQ8 (*31*). Furthermore, we identified the first examples of second-generation cispeptides such as **LVCG**<u>*G*</u>FVNQHLC (Fig. 5C, Figs. S35 and S36), and confirmed the presence of all its precursor peptides in the reaction mixture.

To compare the potential of other cysteine cathepsins to generate cis/transpeptides derived from full-length proteins, we compared the efficacy of hCatS, V, K, L, and B to form

two first-generation cispeptides (**GIVEQCCT**TPKT and **LVCG**GSHLVE). Similar to our previous findings (Fig. 3D), hCatS and hCatV were the most effective proteases, while hCatL, K, and B produced only negligible amounts of these cispeptides (Fig. 5D). These results demonstrate that cis/transpeptidation depends on the specificity of individual cathepsins.

Finally, we used full-length spike and hPF4 proteins to identify cis/transpeptides relevant to thrombotic thrombocytopenia diseases like HIT and VITT. We identified eight reproducibly detected hPF4–spike cis/transpeptides, absent in controls (Fig. 5E, S37). Among these, a cispeptide containing part of an hPF4-VITT epitope (**SLEVIK**LIATLK) (*28*) was of special interest, **SLEVIK** was also observed to fuse with the hPF4-derived nucleophilic peptide in our previous experiment (Fig. 4A, B). Antigenic cispeptides of hPF4 may reveal at least two potential epitopes for antibody binding that may facilitate the antibody interactions with hPF4 tetrameric complexes as described for VITT (*32*).

Identification of cis/transpeptides in cells using CT-TRAP. To investigate cis/transpeptidation in a biologically relevant context, we focused on antigen-presenting cells such as RAW 264.7 macrophages and mouse primary dendritic cells (DCs). 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 (i.e. NH_2-TTSQGGK(N_3)YGRKKRRQRRRK-5-FAM) was designed with a specific architecture: it included an *N*-terminal nucleophilic peptide containing the TTSQ motif, which functions as a competitive substrate for transpeptidation and represents the variable component of the probe that can be adapted to different disease contexts. 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). This reaction yields a Lys-conjugated 4-PEG-Biotin tag that produces a distinct fragment ion (*m/z* 270.1276) in tandem mass spectra (Fig. S40). The fluorogenic 5-FAM motif allowed monitoring of probe internalization, while the TAT-sequence (YGRKKRRQRRRK) enabled efficient lysosomal delivery. Using this probe, we confirmed up to 95% lysosomal uptake in RAW 264.7 macrophages (Fig. 6A; Fig.S39).

Application of CT-TRAP combined with LC-MS/MS on RAW 264.7 macrophages and *de novo* sequencing identified several transpeptides that incorporated the 4-PEG-Biotin tag (Fig. 6B; Fig. S41 and S42), including the partially processed starting material, TTSQGGK(4-PEG-Biotin)YGR. This indicates that the probe was successfully taken up by lysosomes, underwent partial lysosomal processing (YGR↓KK) with an intact Lys-(N_3) handle (allowing post-lysis CuAAC derivatization), and the probe survived the subsequent purification and mass spectrometry preparation steps. None of these peptides were detected in cellular assays lacking the tagged peptide (Fig. 6B). The identification of fragments of the Lys-(4-PEG-Biotin)-containing tag at both the *N*- and *C*-termini of the identified peptides indicated that the nucleophilic 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 cis/transpeptides containing only the Lys-(4-PEG-Biotin) residue within their sequences (Fig. 6B; Fig. S41 and S42), suggesting multi-generation 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 cis/transpeptides was abolished by E-64d, as shown for an individual peptide example (Fig. 6C). This inhibition by E-64d clearly demonstrated that cysteine cathepsins were the proteases responsible for generating these cis/transpeptides within the endolysosomal compartment.

To test whether cis/transpeptidation is an intrinsic feature of antigen processing rather than a disease-associated anomaly, we analyzed insulin fragments in splenocytes and thymic DCs from C57BL/6 mice. Interestingly, the analysis of DCs from mouse splenocytes revealed a transpeptide, **SHLVEALYLVC**FACPKFYTPTK, formed as a hybrid peptide between mouse insulin II and a peptide of bacterial origin (Fig. 6D; Fig. S43). 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

This study resolves the long-standing gap in our knowledge of reverse proteolysis by demonstrating several mechanistic aspects of how cysteine cathepsins catalyze the formation of cis/transpeptides. The use of biotinylated nucleophilic donor peptides provides unambiguous evidence for the formation of cis/transpeptides. The fact that multigenerational transpeptides, first identified with small biotinylated peptides (e.g., **SLEVIK**TTSQGGK-Biotin), were also detected using full-length proteins (e.g., cispeptide **SLEVIK**LIATLK) demonstrates that tagged peptides are powerful tools for transpeptide discovery. The presence of tags such as biotin markedly increases confidence in peptide identification by LC–MS/MS, allowing unambiguous detection of rare cis/transpeptide in complex mixtures. Notably, this strategy enables high-confidence discovery under simplified conditions, while still capturing events that occur in more physiologically relevant protein-protein contexts.

Using this experimental approach, we show that hCatS produces a superagonist peptide previously identified in T1D patients (*5*) and several additional transpeptides with high affinity for T1D-associated HLA-DQ8. Some of them contain fragments from the SARS-CoV-2 spike protein and from human proinsulin. These results underscore the broader role of hCatS in shaping the antigenic repertoire during infection. Importantly, while earlier work had been largely restricted to T1D models, our findings regarding the formation of hPF4-based cis/transpeptides suggest that the same mechanism could also contribute to other autoimmune diseases such as HIT and VITT.

One criticism regarding the relevance of reverse proteolysis lays in the assumption that it may only form negligible, and thus irrelevant, amounts of cis/transpeptides (*6*). In a model reaction using small peptides, one transpeptide accounted for up to 4.5% of one of the hydrolysis products, an efficiency high enough to be biologically relevant and falling within the range of the ~ 2% of transpeptides observed in HLA-DQ8 immunopeptidomes (*5*). Given that a single peptide-MHC complex per cell can trigger a robust T cell response, the levels of cis/transpeptides we observed are likely to be physiologically significant and could initiate an immune response (*33, 34*).

Cathepsin-mediated transpeptidase activity is strongly pH-dependent, with activity increasing when the pH value shifts from 4.5 to 7.5. This is particularly relevant because

lysosomal dysfunction, often accompanied by alkalinization, is a hallmark of several autoimmune diseases (*35*). As the transpeptidase activity is optimal at neutral pH, the extracellular generation of larger transprotein fragments by specific cathepsins appears to be possible, which may then directly act as antigens for B cells. In contrast, the diverse cathepsin content in the endolysosomal compartment may create linear short antigenic peptides suitable for MHC class II presentation.

Also of high relevance is 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 (*36*), suggesting that it may not only break self tolerance, but also accelerate the creation of novel antigenic cis/transpeptides.

The nature of the sequence of cis/transpeptides depends on the specificity of individual cathepsins. As their expression profile differs between the central and the peripheral immune system (*14, 37, 38*), this may lead to the generation of different sets of cis/transpeptides. This could allow self-reactive T cells that escaped negative selection to later encounter these unique cis/transpeptides in the periphery and mount an autoimmune response. For example, it has been shown that increased hCatS expression in peripheral DCs contributes to systemic lupus erythematosus (*39*).

Finally, using our CT-TRAP method allowed for the discovery of highly diverse cis/transpeptides in antigen-presenting cells. We showed that a multigenerational fusion and hydrolysis process by multiple lysosomal proteases actually occurs in cells and substantiates the hypothesis that cis/transpeptidation in antigen-presenting cells may contribute to autoimmunity.

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. This hypothesis calls into question the notion that self-derived peptides which T cells develop a tolerance for remain the same throughout life. Post-translational modification during any time of life can trigger an autoimmune reaction (*40*). Cis/transpeptidation in the context with susceptible autoimmune-associated HLA-haplotypes may represent one of the missing events to

induce autoimmunity. Cis/transpeptides would represent' foreign" antigens for which the immune system is trained for. By linking hCatS to the creation of a known disease-associated neoepitope and other immunogenic cis/transpeptides, 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 investigate the role of reverse proteolysis beyond autoimmunity, including in aging, chronic infection, and cancer, where antigenic remodeling may profoundly influence immune surveillance. It will also be essential to define how cellular conditions, such as lysosomal pH and post-translational modifications, govern the balance between hydrolysis and transpeptidation. Finally, new experimental platforms such as CT-TRAP, together with dedicated bioinformatic approaches, will be required to map cis/transpeptidation events *in vivo* with precision.

References:

1. K. Morihara, Using Proteases in Peptide-Synthesis. *Trends Biotechnol* **5**, 164-170 (1987).
2. 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).
3. M. Mishto *et al.*, Driving Forces of Proteasome-catalyzed Peptide Splicing in Yeast and Humans. *Molecular & Cellular Proteomics* **11**, 1008-1023 (2012).
4. P. Faridi *et al.*, A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci Immunol* **3**, (2018).
5. M. T. Tran *et al.*, T cell receptor recognition of hybrid insulin peptides bound to HLA-DQ8. *Nature communications* **12**, 5110 (2021).
6. A. Admon, Are There Indeed Spliced Peptides in the Immunopeptidome? *Molecular & Cellular Proteomics* **20**, (2021).
7. J. A. Villadangos *et al.*, Proteases involved in MHC class II antigen presentation. *Immunol Rev* **172**, 109-120. (1999).

8. P. R. Nielsen, T. W. Kragstrup, B. W. Deleuran, M. E. Benros, Infections as risk factor for autoimmune diseases - A nationwide study. *J Autoimmun* **74**, 176-181 (2016).
9. M. K. Smatti *et al.*, Viruses and Autoimmunity: A Review on the Potential Interaction and Molecular Mechanisms. *Viruses-Basel* **11**, (2019).
10. D. B. Cines, J. B. Bussel, SARS-CoV-2 Vaccine-Induced Immune Thrombotic Thrombocytopenia. *N Engl J Med* **384**, 2254-2256 (2021).
11. C. Sharma, J. Bayry, High risk of autoimmune diseases after COVID-19. *Nat Rev Rheumatol* **19**, 399-400 (2023).
12. K. Saegusa *et al.*, Cathepsin S inhibitor prevents autoantigen presentation and autoimmunity. *J Clin Invest* **110**, 361-369 (2002).
13. R. J. Riese *et al.*, Essential role for cathepsin S in MHC class II-associated invariant chain processing and peptide loading. *Immunity* **4**, 357-366 (1996).
14. C. Stoeckle *et al.*, Cathepsin S dominates autoantigen processing in human thymic dendritic cells. *J Autoimmun* **38**, 332-343 (2012).
15. Y. Wang *et al.*, Nucleophile labeling of cysteine and serine protease substrates. *J Biol Chem* **271**, 28399-28404 (1996).
16. E. Olson, M. Raghavan, Major histocompatibility complex class I assembly within endolysosomal pathways. *Current Opinion in Immunology* **84**, (2023).
17. J. Wang, J. Miao, P. Zhu, Insights into the complexities of Citrullination: From immune regulation to autoimmune disease. *Autoimmun Rev* **24**, 103734 (2025).
18. K. N. Henderson *et al.*, A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. *Immunity* **27**, 23-34 (2007).
19. Y. Wang *et al.*, How C-terminal additions to insulin B-chain fragments create superagonists for T cells in mouse and human type 1 diabetes. *Sci Immunol* **4**, (2019).
20. S. I. Mannering, A. F. Rubin, R. Wang, P. Bhattacharjee, Identifying New Hybrid Insulin Peptides (HIPs) in Type 1 Diabetes. *Front Immunol* **12**, 667870 (2021).
21. B. Reed *et al.*, Lysosomal cathepsin creates chimeric epitopes for diabetogenic CD4 T cells via transpeptidation. *Journal of Experimental Medicine* **218**, (2021).
22. O. Molberg *et al.*, Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat Med* **4**, 713-717 (1998).

23. B. I. Aydogan, U. Unluturk, M. Cesur, Type 1 diabetes mellitus following SARS-CoV-2 mRNA vaccination. *Endocrine* **78**, 42-46 (2022).
24. A. S. Alsudais *et al.*, Autoimmune diabetes mellitus after COVID-19 vaccination in adult population: a systematic review of case reports. *BMC Endocr Disord* **23**, 164 (2023).
25. C. Debuysschere, M. P. Nekoua, E. K. Alidjinou, D. Hober, The relationship between SARS-CoV-2 infection and type 1 diabetes mellitus. *Nature reviews. Endocrinology* **20**, 588-599 (2024).
26. A. Huynh *et al.*, Characteristics of VITT antibodies in patients vaccinated with Ad26.COV2.S. *Blood Adv* **7**, 246-250 (2023).
27. A. Huynh, J. G. Kelton, D. M. Arnold, M. Daka, I. Nazy, Antibody epitopes in vaccine-induced immune thrombotic thrombocytopaenia. *Nature* **596**, 565-+ (2021).
28. E. Petito *et al.*, 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).
29. A. Mittal, A. Khattri, V. Verma, Structural and antigenic variations in the spike protein of emerging SARS-CoV-2 variants. *PLoS Pathog* **18**, e1010260 (2022).
30. S. Kaabinejadian *et al.*, Accurate MHC Motif Deconvolution of Immunopeptidomics Data Reveals a Significant Contribution of DRB3, 4 and 5 to the Total DR Immunopeptidome. *Front Immunol* **13**, 835454 (2022).
31. 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. *Journal of Clinical Investigation* **115**, 2268-2276 (2005).
32. V. A. S. Dabbiru, L. Muller, L. Schonborn, A. Greinacher, Vaccine-Induced Immune Thrombocytopenia and Thrombosis (VITT)-Insights from Clinical Cases, In Vitro Studies and Murine Models. *J Clin Med* **12**, (2023).
33. J. Huang *et al.*, A Single Peptide-Major Histocompatibility Complex Ligand Triggers Digital Cytokine Secretion in CD4 T Cells. *Immunity* **39**, 846-857 (2013).
34. R. Platzer *et al.*, Monomeric agonist peptide/MHCII complexes activate T-cells in an autonomous fashion. *Embo Rep* **24**, (2023).

35. W. Zhang *et al.*, Defective Lamtor5 Leads to Autoimmunity by Deregulating v-ATPase and Lysosomal Acidification. *Adv Sci (Weinh)* **11**, e2400446 (2024).
36. A. M. Curran *et al.*, Citrullination modulates antigen processing and presentation by revealing cryptic epitopes in rheumatoid arthritis. *Nature communications* **14**, 1061 (2023).
37. E. Tolosa *et al.*, Cathepsin V is involved in the degradation of invariant chain in human thymus and is overexpressed in myasthenia gravis. *J Clin Invest* **112**, 517-526 (2003).
38. 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. *Front Immunol* **11**, 453 (2020).
39. S. J. Kim *et al.*, Increased cathepsin S in Prdm1(-/-) dendritic cells alters the T(FH) cell repertoire and contributes to lupus. *Nat Immunol* **18**, 1016-1024 (2017).
40. T. Mustelin, F. Andrade, Autoimmunity: the neoantigen hypothesis. *Front Immunol* **15**, (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-TTSQGGK(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 cispeptides (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: Nucleophilic peptide specificity and citrullination effects on the cis/transpeptidase activity of human cathepsin S. (A) Schematic of the method used to detect biotinylated cis/transpeptides generated by cathepsins. (B) SDS-PAGE of Coomassie-positive hydrolysis and cis/transpeptides and Avidin-FITC detected cis/transpeptides derived from bovine serum albumin (BSA) incubated with hCatS in the presence of three different biotinylated nucleophilic peptides at pH 6.5 for 2 h at RT, as well as in the presence or absence of the cathepsin inhibitor E-64. (C) Quantification of FITC detectable biotinylated cis/transpeptides in the presence of three different nucleophilic peptides. * Biotinylated peptide was added after BSA hydrolysis and CatS inhibition to exclude non-specific nucleophilic peptide binding. (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 transpeptide generation from Ac-AE(Ed)F*Cit*TTSQGGK(Dab) by hCatS. (F,G) Time-course formation of transpeptides in independent reactions with citrullinated or non-citrullinated substrates. (H,I) Corresponding hydrolysis products from reactions in panels C and E. (J) Transpeptide formation rates in panels F-I, calculated as peak area divided by time. (Bio: biotin; Dab: Dabcyl; Ed: Edans).

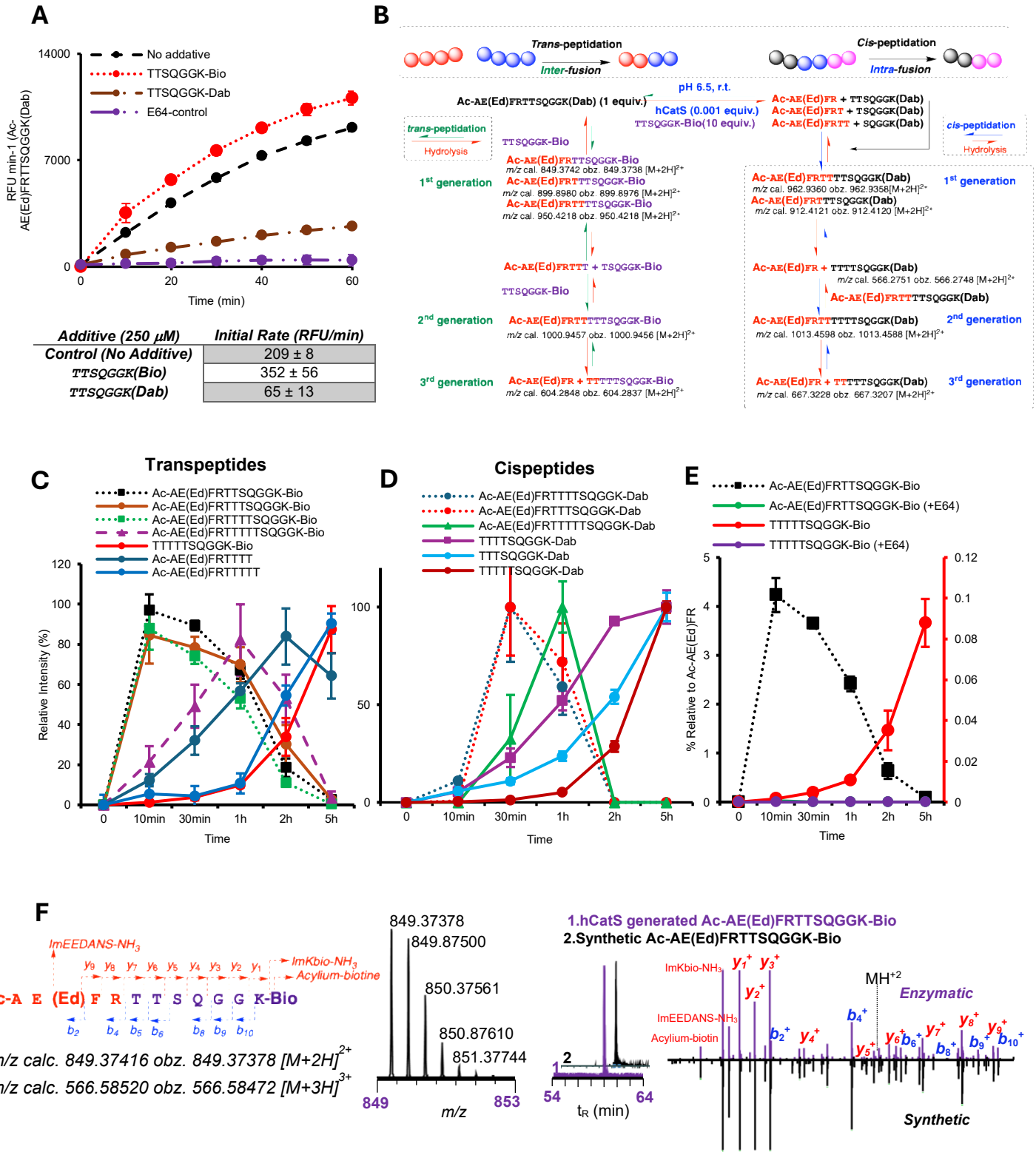
Fig. 3. Generation of putative TID super-agonists by cysteine cathepsins. (A) Selected biotinylated transpeptides identified in a reaction mixture containing proinsulin and NH_2-NAVEGGK-Biotin using LC-MS/MS and the location splicing site in a then proinsulin sequence (in red). (B) Chromatogram comparing retention times of ELGGGNAVEGGK-Bio generated in reactions with hCatS in the absence or presence of E-64. 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 E-64 inhibition. F) HLA-DQ8 binding assay (performed by ProImmune Ltd.) compares the binding capacity of 15-mer versions of discovered transpeptides. (Bio: biotin).

Fig. 4. Generation of cis/transpeptides by hCatS in the presence of hPF4 or spike protein subunits and a hPF4 derived antigenic biotinylated nucleophilic 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 hPF4 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 E-64. (C) Proposed mechanistic pathway for the formation of the identified third-generation transpeptide (TTSTTTTTTSQGGK-Bio). (D) Avidin-FITC blot of the reaction mixtures of S1 and S2 subunits (S1 and S2) incubated with hCatS and NH_2-TTSQGGK-Bio for 2 h at RT at different pH values. (E) Quantification of Avidin-FITC signals from (D). (F) Biotinylated transpeptides identified by LC-MS/MS from reaction mixtures containing S1 and S2 subunits, NH_2-TTSQGGK-Biotin, and hCatS at pH 6.5 for 2 h. (Bio: biotin).

Fig. 5. Cis/transpeptides generated by hcatS and other cysteine cathepsins from full-length proinsulin, hPF4, and S1/S2 subunit proteins. (A) Reaction mixtures containing

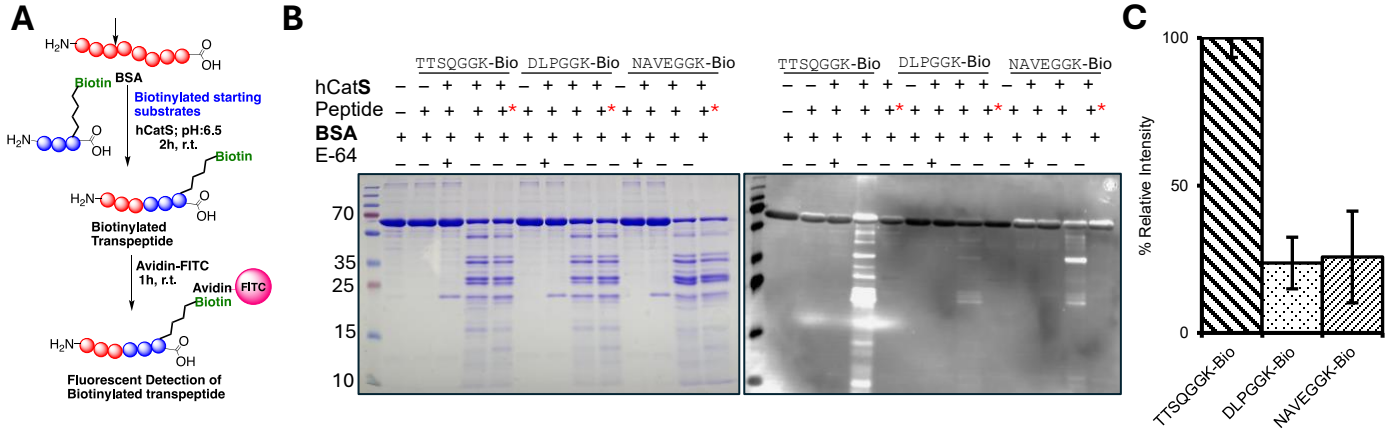
SARS-CoV-2 spike protein, human proinsulin, and hCatS at pH 6.5 for 2 h at RT were analyzed in triplicates. Cis- and transpeptides were reproducibly detected in all three experimental replicates, whereas no corresponding products were observed in controls. Peptides were identified by LC–MS/MS de novo sequencing with ALC% (Average Local Confidence) ≥ 90, ensuring high-confidence assignments. (B) Selected first-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 E-64. (D) Selected first- and second-generation cispeptides derived from proinsulin by hCatS and their proposed pathway of formation. (E) Comparing the formation of two selected cispeptides by different cysteine cathepsins. (F) First-generation of cis/transpeptides identified from a reaction mixture containing spike proteins, hPF4, and hCatS at pH 6.5 for 2 h at RT.

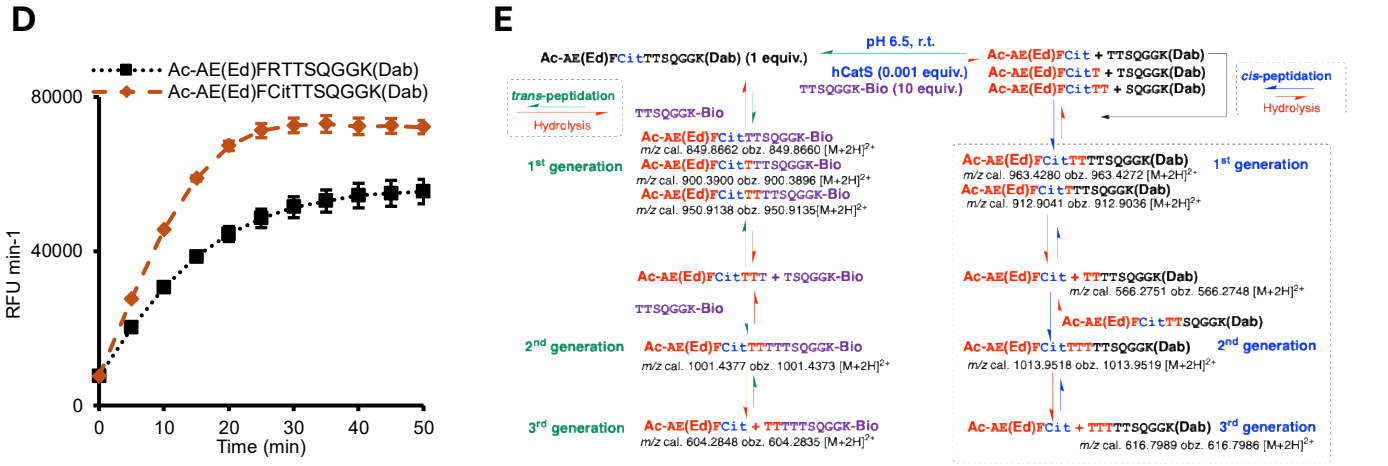
Fig. 6. Identification of cis/transpeptides in antigen-presenting cells (APCs). (A) Scheme of the CT-TRAP method using a lysosomotropic TAT peptide (NH_2-TTSQGGK(N_3)-TAT-5FAM) and its cellular uptake to identify cis/transpeptides in cells. This method incorporated the inclusion of highly specific signature ions allowing the unequivocal identification of TAT-containing cis/transpeptides. (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 cis/transpeptides. (C) Ion chromatograms of a selected cis/transpeptides 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, E-64d. (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).



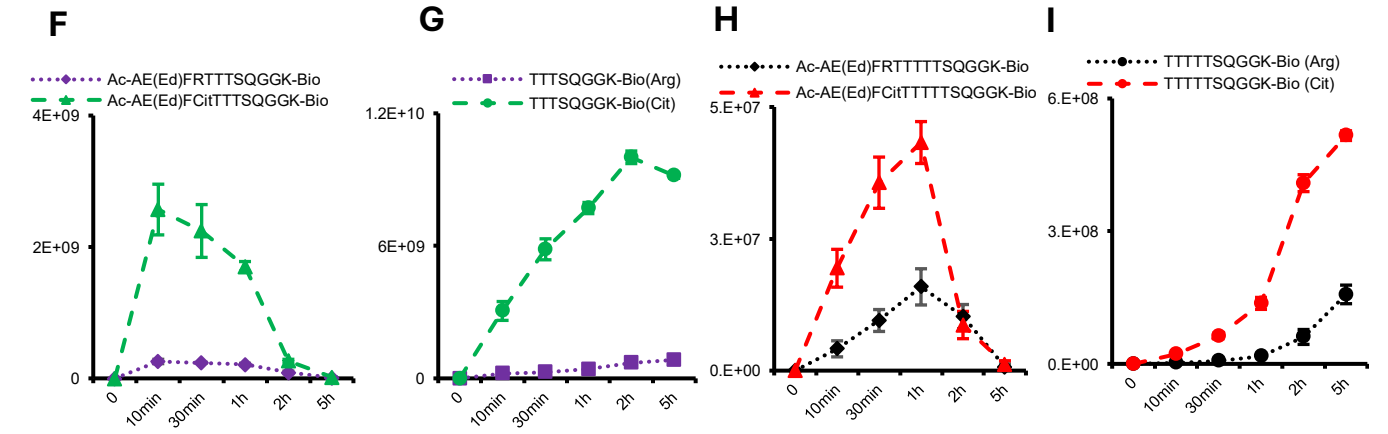
Additive (250 μM)	Initial Rate (RFU/min)
Control (No Additive)	209 ± 8
TTSQGGK(Bio)	352 ± 56
TTSQGGK(Dab)	65 ± 13

Figure 1





Substrate	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\mu M^{-1}\ s^{-1}$)
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

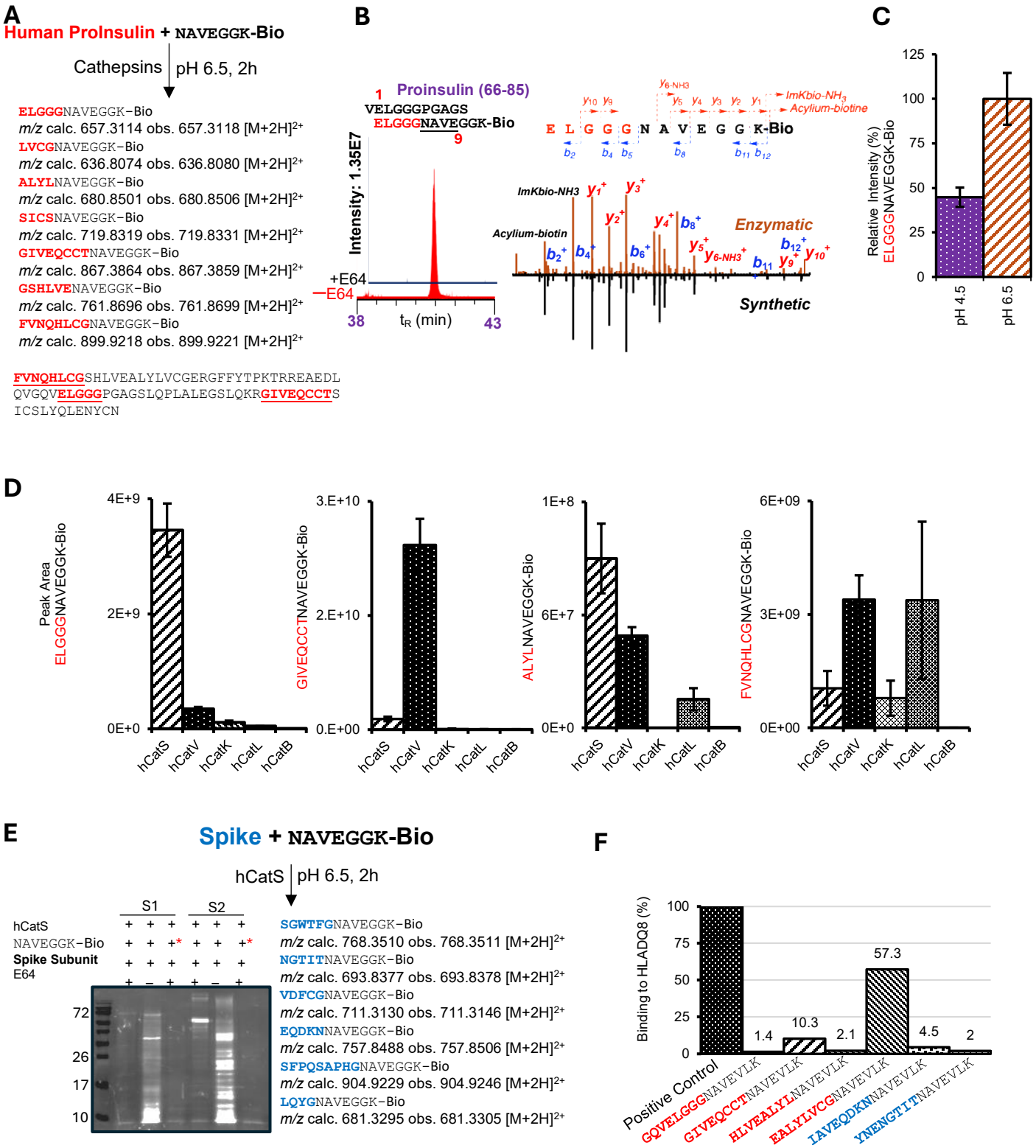

A
Human ProInsulin + NAVEGGK-Bio
Cathepsins
pH 6.5, 2h
ELGGGNAVEGGK-Bio
m/z calc. 657.3114 obs. 657.3118 [M+2H]2+
LVCGNAVEGGK-Bio
m/z calc. 636.8074 obs. 636.8080 [M+2H]2+
ALYLNAVEGGK-Bio
m/z calc. 680.8501 obs. 680.8506 [M+2H]2+
SICSNAVEGGK-Bio
m/z calc. 719.8319 obs. 719.8331 [M+2H]2+
GIVEQCCTNAVEGGK-Bio
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+
FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDL
QVGQVELGGGPGAGSLQPLALEGSLQKRGIVEQCCTS
ICSLYQLENYCN
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
72
26
17
10
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 (%)
1.4
10.3
2.1
57.3
4.5
2
Positive Control
GQVELGGGNAVEVLK
GIVEQCCTNAVEVLK
HLVEALYLNAVEVLK
EALYLVCGNAVEVLK
IAVEQDKNNAVEVLK
YNENGTITNAVEVLK

Figure 3

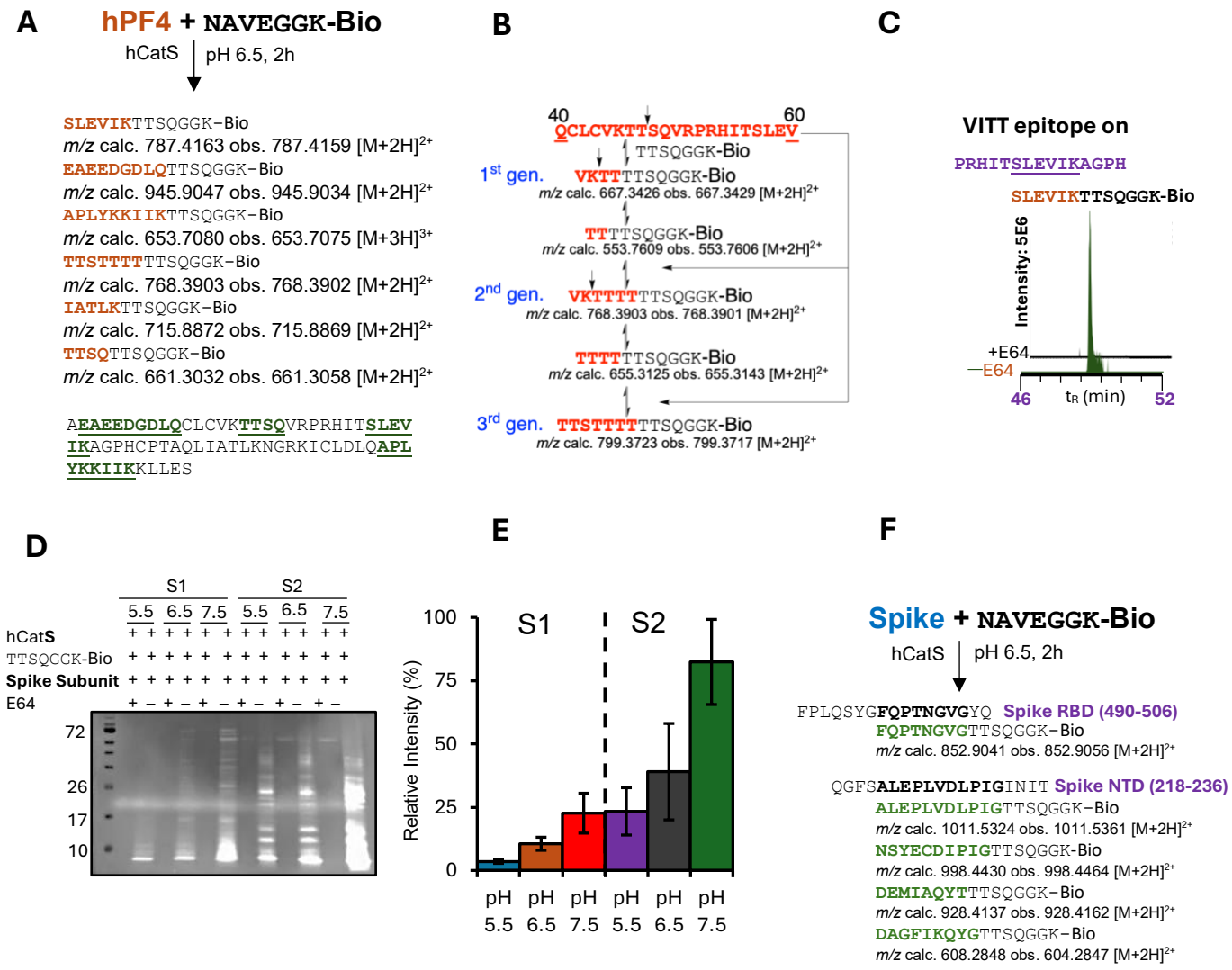

A
hPF4 + NAVEGGK-Bio
hCatS
pH 6.5, 2h
SLEVIKTTSQGGK-Bio
m/z calc. 787.4163 obs. 787.4159 [M+2H]2+
EAEEDGDLQTTSQGGK-Bio
m/z calc. 945.9047 obs. 945.9034 [M+2H]2+
APLYKKIIKTTSQGGK-Bio
m/z calc. 653.7080 obs. 653.7075 [M+3H]3+
TTSTTTTTTSQGGK-Bio
m/z calc. 768.3903 obs. 768.3902 [M+2H]2+
IATLKTTSQGGK-Bio
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

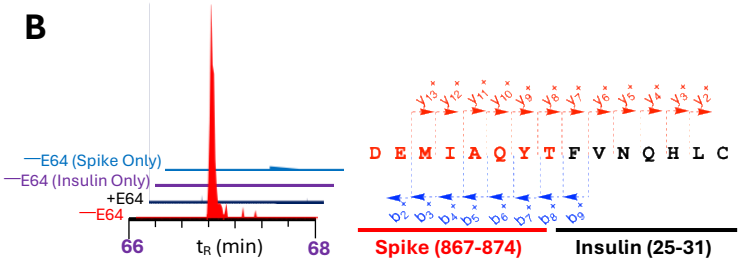
A

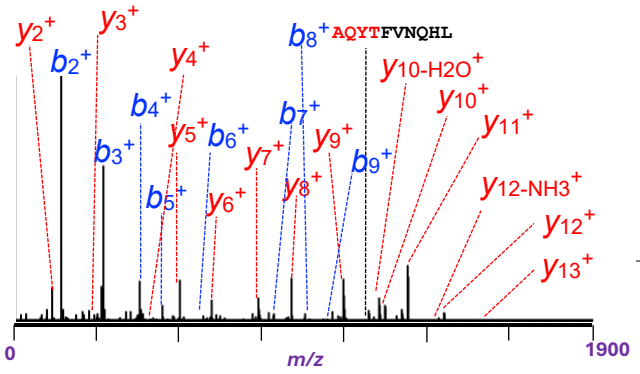
Spike + Human Proinsulin —hCatS, pH 6.5, 2h→ Cis/transpeptides
Spike-Insulin = 23
Insulin-Insulin = 45

Spike	Insulin	1st gen. Transpeptide	ALC%
699-LGAENSVA-706	33-SHLVE-37	LGAENSVASHLVE	94.6
701-AENSVA-706	33-SHLVE-37	AENSVASHLVE	94.7
837-YGDCLG-842	25-FVNQHLC-31	YGDCLGFVNQHLC	94.3
867-DEMIAQYT-874	25-FVNQHLC-31	DEMIAQYTFVNQHLC	96.6
613-QDVNCT-618	50-YTPKT-54	QDVNCTYTPKT	96.3
867-DEMIAQYT-874	33-SHLVE-37	DEMIAQYTSHLVE	95
867-DEMIAQYT-874	50-YTPKT-54	DEMIAQYTYTPKT	92
867-DEMIAQYT-874	45-ERGFFY-50	DEMIAQYTERGFFY	95.2
797-FGGFNFS-803	33-SHLVE-37	FGGFNFSSHLVE	94

Insulin	Insulin	1st gen. Cispeptide	ALC%
41-LVCG-44	33-SHLVE-37	LVCGGSHLVE	98.4
33-SHLVE-37	25-FVNQHLC-31	SHLVEFVNQHLC	98.2
41-LVCG-44	25-FVNQHLCG-32	LVCGFVNQHLCG	97.2
101-SLYQLE-106	25-FVNQHLC-31	SLYQLEFVNQHLC	92.2
101-SLYQLE-106	50-YTPKT-54	SLYQLEYTPKT	91.1
97-TSICS-101	32-GSHLVE-37	TSICSGSHLVE	96.7
90-GIVEQCCT-97	51-TPKT-54	GIVEQCCTTPKT	94.1
41-LVCG-44	48-FFYTPKT-54	LVCGFFYTPKT	94.7

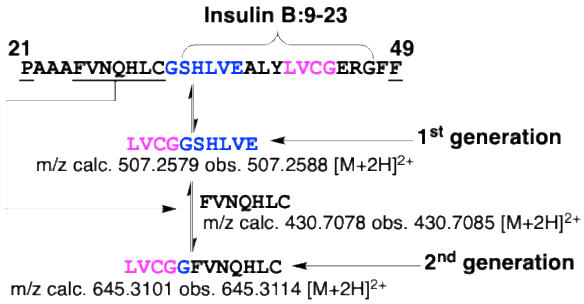
B



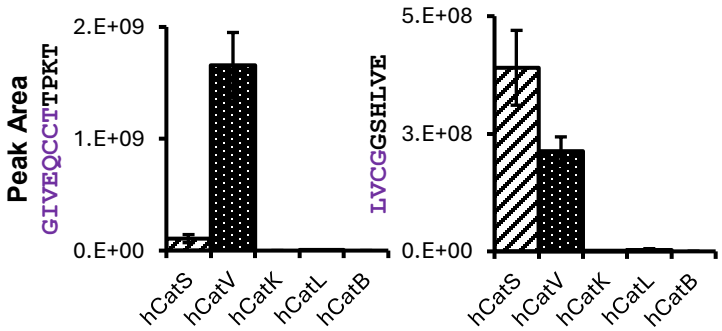


C

1st gen. Cispeptide	Insulin	2nd gen. Cispeptide
LVCGGSHLVE	AFVNQHLCGSHL	LVCGGFVNQHLC
GIVEYTPKT	ALYLVCGERGF	VCGGIVEYTPKT
LVCGTPKT	CGSHLVEALYL	SHLVEGTPKT
LVCGTPKT	SICSLYQLENY	SLYQLEGTPKT
LVCGFVNQHLC	GFFYTPKTRREA	VCGFVNTPKT



D



E

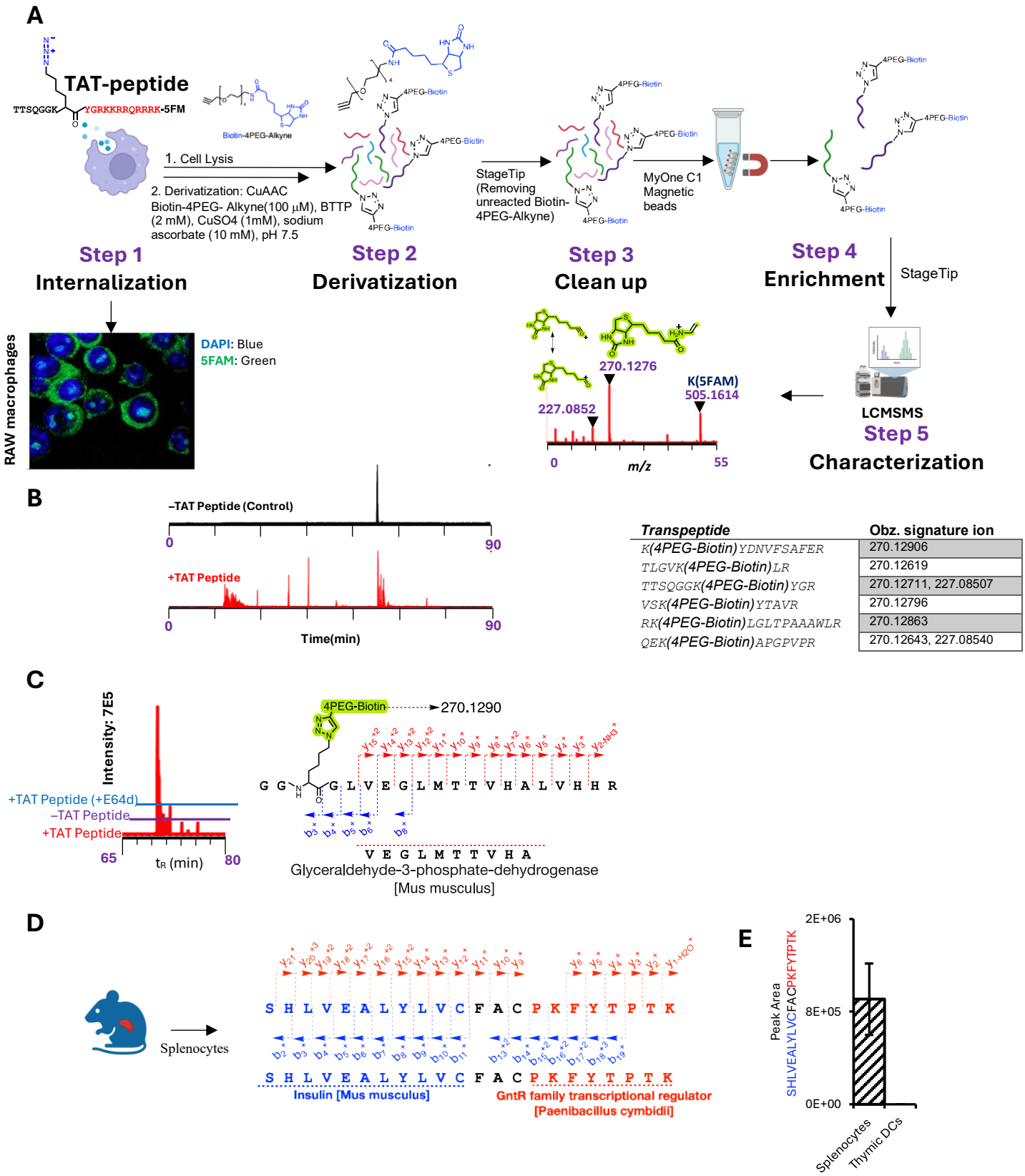
Spike + Human PF4 —hCatS, pH 6.5, 2h→ Cis/transpeptides

Spike	hPF4	1st gen. Transpeptide	ALC%
230-LPIG-233	72-LIATLK-77	TAGALIATLK	95.2
877-LLAG-880	65-PHCPTAQ-71	LLAGPHCPTAQ	97.7

hPF4	hPF4	1st gen. Cispeptide	ALC%
57-SLEVIK-62	72-LIATLK-77	SLEVIKLIATLK	94.7
57-SLEVIK-62	88-APLYK-92	SLEVIKAPLYK	96.7
57-SLEVIK-62	32-EAEEDGDLQCLCVK-45	SLEVIKEAEEDGDLQCLCVK	97.1
32-EAEEDGDLQ-40	88-APLYK-92	EAEEDGDLQAPLYK	96
32-EAEEDGDL-39	44-CVKTTS-49	EAEEDGDLCVKTTS	95.2
32-EAEEDGDLQ-40	63-AGPHCPTAQ-71	EAEEDGDLQAGPHCPTAQ	93.9

Figure 5

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



Transpeptide	Obz. signature ion
K(4PEG-Biotin)YDNVFSAFER	270.12906
TLGVK(4PEG-Biotin)LR	270.12619
TTSQGGK(4PEG-Biotin)YGR	270.12711, 227.08507
VSK(4PEG-Biotin)YTAVR	270.12796
RK(4PEG-Biotin)LGLTPAAAWLR	270.12863
QEK(4PEG-Biotin)APGPVPR	270.12643, 227.08540

Figure 6