A protein microarray-based *in vitro* transglutaminase assay platform for epitope mapping and peptide discovery: implication for transglutaminase-mediated immunodominant determination

Chen Liu[1,2]*

[1]Bellaire Foundation and University of Houston, [2]Department of Biology and Biochemistry, University of Houston, and * correspondence to Chen Liu (cliu8@uh.edu).

Abstract

Transglutaminases (TGs) are a family of crosslinking enzymes catalyzing the formation of intra- and inter-molecular glutamine-lysine isopeptide bonds in a calcium dependent manner. Activation of transglutaminases is pathogenically associated with severe human diseases including neurodegenerations, cardiovascular diseases, and autoimmune diseases. Although continuous efforts determining the enzymes' substrate preference have been witnessed, a high-throughput assay platform with the "omic" efficiency is still missing for the global identification of substrate-specific TG modification sites. In this study we report a protein microarray-based *in vitro* TG assay platform for rapid and large-scale (up to 30000 reactions per chip) determination of the glutamine (Q)-bearing TG modification motifs. With this platform we identified the Q16 in superoxide dismutase 1 and Q109 in alpha-synuclein as the primary modification sites for tissue transglutaminase (TG2), the most ubiquitous member of the enzyme family. Of particular interest, we found a close match between the modification motifs and published vaccine epitope sequences in alpha-synuclein, implying an essential and intrinsic role transglutaminase might play in the determination of immunodominant epitopes. Our data collectively suggest the glutamine and its follow-up five residues on the C terminal of a protein compose a minimal determinant motif for TG2 modification and the TG2 modification motifs determined by our platform could finally correspond to the substrate's epitope sequences in antigen processing. To screen for site-specific interfering peptides and assist gene editing-based protein engineering, we optimized through *onchip* amino-acid scanning the TG2 modification motif QQIV in the extracellular matrix protein fibronectin and

obtained 14 variants with significantly higher TG2 reactivity that might serve as the competitive inhibitor peptides and 1 with lower reactivity. We further confirmed this optimization approach with 12-mer peptides, the longest ones that could be synthesized on the chip. Taken together, our synthetic transglutaminase assay platform might be able to deliver a precise epitope blueprint for immunotherapeutic targeting and provide proof-of-concept and directional studies for TG-based peptide discovery and protein design.

Introduction

Transglutaminases (TGs) in mammalian cells are a family of crosslinking enzymes posttranslationally catalyzing the covalent isopeptide bond formation between the γ-carboxamide groups of a peptide-bound glutamine residue and a free amine or a peptide-bound lysine in a calcium dependent manner[1-4]. Functionally, TG-mediated isopeptide modification results in the stabilization and aggregation of substrate proteins by facilitating the assembly of supramolecular structure resistant to proteolysis[1, 5]. Activation of transglutaminase and TG-mediated isopeptide modification have been demonstrated to be pathogenically associated with severe human diseases including neurodegenerations[6-9], cardiovascular diseases[10-13], and autoimmune diseases[12, 14, 15]. Accumulation of pathogenic proteins crosslinked by transglutaminase has long been recognized as a hallmark of major neurodegenerative disorders including Alzheimer's disease (AD)[16-21], Parkinson's disease (PD)[22, 23], and Huntington's disease (HD)[24, 25]. In cardiovascular diseases, transglutaminases contribute to the formation of atherosclerotic plaques by crosslinking extracellular matrix (ECM) proteins[26-31] and to hypertensive disorders by modifying and sensitizing vasopressor receptors[32, 33]. Transglutaminases are also well-known to participate in autoimmune responses through posttranslational modifications favoring the generation of neoantigens[34-37]. Successful amelioration of key disease features by transglutaminase inhibitors in the pre-clinical animal models of related disorders not only establishes TG as an effective therapeutic target but also potentiates transglutaminase inhibitors as useful drugs for the disease treatment[38-42]. However, diagnosis of the related complications in patients

usually accompanies a significant build-up of TG-crosslinked aggregates or plaques, and simply inhibiting the enzymes' activity is not sufficient to clean the build-ups. Given this concern, characterizing substrate-specific crosslinking sites and understanding the sequence preference of transglutaminase substrate would be of great importance in designing immunotherapeutic strategies capable of cleaning the aggregates to revoke the disease progression.

To determine the enzymes' substrate preference and map the modification sites among various protein substrates, continued efforts using phage display[43-45], mass spectrometry[46-49], protein arrays[50, 51], and bioinformatics tools[52-54] have been witnessed. Although these studies greatly advanced our knowledge regarding the modification patterns of transglutaminases, a high-throughput platform with the efficiency of systems biology is still missing for the global identification of substrate-specific TG modification sites. In this study, by combining the tagged amine donor dansyl-cadaverine-based *in vitro* TG assay[55] with a protein microarray[56] we report a platform for rapid and large-scale (up to 30000 reactions per chip) determination of the glutamine (Q)-bearing TG modification motifs. We tested the platform with peptides from neurodegenerative proteins and mapped the primary modification sites for tissue transglutaminase (TG2), the most ubiquitous member of the enzyme family[1-4], in superoxide dismutase 1 and alpha-synuclein. Our data indicate the glutamine and its follow-up five residues on its C terminal compose a minimal determinant motif for TG2 modification. To manipulate the TG2 modifications on a certain protein and screen for site-specific interfering peptides, we

employed an *onchip* amino-acid scanning[57, 58] method for the optimization of modification motifs. By scanning the TG2 modification motif QQIV in the extracellular matrix protein fibronectin and 12-mer Q-bearing peptides, the longest ones that could be synthesized on the chip, we confirmed the approach's capability to serve peptide discovery and protein engineering.

Methods

Onchip in vitro TG2 assay

To identify the glutamine residues that can be modified by tissue transglutaminase on the peptide microchip, the synthesized peptide microchip was incubated with 100 ug/ml guinea pig liver tissue transglutaminase (Sigma) and 3 mM dansyl-cadaverine (Sigma) in 1 ml of TBS containing 5 mM Calcium Chloride and 1 mM DTT at 37 degree for 30 minutes. Afterwards, the peptide chip was washed at least 3 times with TBS. After washing off tissue transglutaminase and dansyl-cadaverine molecules bound on the synthesized peptides, the dansyl-cadaverine conjugated on the chip was tracked by rabbit anti-dansyl antibody (Invitrogen) followed by Alexa Fluor 594-labeled anti-rabbit secondary antibody (Invitrogen). Fluorescent microchip figures were quantified and analyzed with ArrayPro32. Original array figures and data are available upon request.

Results

Mapping TG2 modification sites in neurodegenerative proteins with high-throughput *in vitro* assay platform

To establish a high-throughput assay platform for the rapid and large-scale

identification of TG2 modification sites in disease-related proteins, we synthesized on microchips the glutamine-bearing motifs in superoxide dismutase 1 (SOD1) and alpha-synuclein, the pathogenic proteins in amyotrophic lateral sclerosis (ALS) and Parkinson's disease, respectively.

To characterize the glutamine-bearing TG2 modification motifs on these synthesized peptides, dansyl-cadaverine, a well-established amine donor in transglutaminase reaction, was covalently conjugated to the glutamine residues on the peptide chip by purified TG2 with the help of calcium (Figure 1A). After washing off tissue transglutaminase and dansyl-cadaverine molecules bound on the synthesized peptides, the dansyl-cadaverine conjugated on the chip was probed by anti-dansyl antibody followed by Alexa Fluor 594-labeled anti-rabbit secondary antibody. In this way the level of dansyl-cadaverine incorporation on a certain peptide was measured by the fluorescent intensity.

Each peptide synthesized on the chip has a length of at least 4 amino acids, and its maximum length could be up to 12 mer. And the glutamine residue needs to appear in each position of the peptide once to ensure the thorough coverage of the screening (Figure 1B). In this way the peptide screening may also pattern the substrate sequence optimal for the modification. For example, in human alpha-synuclein the surrounding sequence of the first glutamine is AEKTKQ$_{24}$GVAEAA (Figure 1B). So the sequences for its 4 mer peptides would be QGVA, KQGV, TKQG, and KTKQ, and those for 5 mer would be QGVAE, KQGVA, TKQGV, KTKQG, and EKTKQ. Up to 11 residues on either N

or C side of the glutamine is covered on the chip. Each peptide is named as *protein name_Q residue# in the protein_N terminal residue# in the peptide~C terminal residue# in the peptide (In the peptide position of Q is defined as 0, and the one directly before or after it as -1 or 1)*. So the 4 mer peptide $KQ_{24}GV$ is named as α-synuclein_Q24_-1~2 (Figure 1B). Therefore, for each glutamine residue in a given protein, the initial number of peptide variants synthesized on the chip will be 4+5+6+7+8+9+10+11+12=72. The corresponding peptides with Q to S swap are also synthesized on the same chip as negative controls.

With this approach we identified Q16 in superoxide dismutase 1 and Q109 in alpha-synuclein as the primary modification sites for TG2 (Figure 2B and 2C). In human SOD1 protein there are 3 glutamine residues including Q16, Q23, and Q154. Compared with Q23 and Q154 counterparts, the Q16 peptides with the 5 mer motif directly following the glutamine residue (QGIINF) showed significantly higher fluorescent intensities (>6000), and the Q to S swap could effectively reduce their fluorescent levels, indicating the Q16 residue is the most probable TG2 modification site in the protein (Figure 2B). Similarly, among the 6 glutamine residues in human alpha-synuclein protein, the Q109 residue with its follow-up 5 mer motif (QEGILE) elicited the strongest dansyl-cadaverine incorporation signal and thereby was identified as the TG2 modification site of the protein (Figure 2C). Aligned data with top hits from SOD1 and alpha-synuclein collectively suggest that the Q and its follow-up 5 residues compose a minimal determinant motif for TG2 modification (Figure 2D), which is further confirmed by the peptides with truncated minimal determinant motifs (not shown).

Optimizing TG2 modification sites with the *onchip* amino-acid scanning

The small peptide QQIV is a transglutaminase substrate identified in the extracellular matrix protein fibronectin. As an amine acceptor, the peptide has been demonstrated to be an effective competitive inhibitor for the transglutaminase reaction[59]. Given this, we chose it as one of the positive substrate peptides on the peptide chip. On our assayed chip the QQIV peptide showed a fluorescent intensity of ~5000 which is much higher than those (~800) of four-residue negative control peptides without glutamine residues and the chip background (Figure 3A). These results confirmed the efficiency and specificity of the dansyl-cadaverine incorporation in our assay system. To scan for a more preferred TG2 substrate and thereby for a more optimized competitive inhibitor than QQIV, we swapped the isoleucine (I) residue in this small peptide to every other amino acid. Except one variant with reduced transglutaminase reactivity, 14 out of the 19 swapped peptides show significantly higher fluorescent signals and the improvement could be up to 3 folds (Figure 3A), suggesting their better candidacy for competitive inhibitor and substrate of TG2 modification than the original QQIV. However, transglutaminase reactivity of most of these 14 mutants was significantly compromised when the hydrophobic V residue was changed to the hydrophilic G (data not shown). To further test the amino acid-scanning approach with the minimal determinant motif for TG2 modification, we swapped to any other amino acid the first leucine residue in the 12-mer small peptide sequence REQLYLDYNVFS, a known TG2 substrate found in a phage display library. Through pan-amino acid scanning at the L residue we found 6 variants (L to N,

S, E, R, V or T) with significantly higher transglutaminase reactivity and 6 (L to G, K, M, W, Y or F) with lower reactivity (Figure 3B). Taken together, our result indicates that the residues within the minimal determinant motif could be reasonable targets for the optimization of modification site and the design of substrate-specific interfering peptides.

Elevating peptide's reactivity with transglutaminase by adding glutamine repeats

Characterized in neurodegenerative complications like the polyglutamine diseases, glutamine repeats elicit excellent TG substrate properties in polypeptides by functioning as efficient amine acceptors, and thereby are considered as a biochemical cause for the pathogenesis. To test this feature on our platform, we randomly added the double glutamine repeat motif QQXX (X stands for any of the 20 amino acids) at the N terminal of the peptide QQIV. Consistent with previous findings, the addition of double glutamine repeat motif resulted in a significant upregulation of the TG2-mediated conjugation of dansyl-cadaverine as the average fluorescent intensity (~10000) of the 400 QQXXQQIV peptides is two folds higher than that (~5000) of the original QQIV peptide (Figure 4A) and 297 out of the 400 QQXXQQIV peptides show a fluorescent intensity of more than 5000 on our assayed chip (Figure 4B). With an average fluorescent intensity of ~20000 (Figure 4A), the top 40 QQXXQQIV peptides could serve as competitive inhibitors to block TG2's modification on the QQIV motif of fibronectin. And among them, the peptides with P at the third residue or I at the fourth show up at most times (Figure 4C), indicating a preferred pattern for the linker region between glutamine repeats.

Discussion

In this study we report a protein microarray-based *in vitro* TG assay platform for fast and high throughput identification of the glutamine (Q)-bearing TG modification motifs among proteins of interest. With this platform we characterized the Q16 in superoxide dismutase 1 and Q109 in alpha-synuclein as the primary TG2 modification sites. Our data indicate the glutamine and its follow-up five residues make up a minimal determinant motif for TG2 modification. To optimize the site for TG2 modification we used an *onchip* amino-acid scanning method in which the residues near the glutamine residue are swapped with any other amino acid to manipulate its modification level. Using this approach we scanned the isoleucine residue in TG2 modification motif QQIV of fibronectin and obtained 14 variants with significantly higher TG2 reactivity and 1 with lower reactivity. We further confirmed this approach by scanning the first leucine residue within the minimal determinant motif of the 12-mer peptide REQLYLDYNVFS. Consistent with the findings from polyglutamine diseases, we also showed that the addition of glutamine repeats with optimized linker residues are able to significantly elevate the peptide QQIV's transglutaminase reactivity.

Protein aggregation is a hallmark of neurodegenerative diseases including but not limited to Alzheimer's disease, Huntington's disease, and Parkinson's disease, where TG2-mediated crosslinking of the aggregates has been consistently characterized. Moreover, by preventing substrates'

ubiquitination[60], TG2 crosslinking could directly impair the targeted degradation of the aggregates and is thus considered to play a pathogenic role. To map the crosslinking sites for the development of immunotherapies targeting the pathogenic aggregates, we designed a protein microarray-based *in vitro* TG2 assay platform and confirmed its efficacy by identifying the TG2 modification sites in alpha-synuclein and superoxide dismutase 1. With this approach the Q109 in alpha-synuclein is characterized as the primary glutamine site for the enzyme modification. Other residues near the c-terminus including Q79 and Q134 are also identified as modification sites. Of particular interest, these modification motifs correspond to the epitope sequences found in animals with full-length protein immunization[61, 62], in which antibodies recognize aa85-99, aa109-123, aa112-126, and aa126-138; B cells aa106-125; and T cells aa76-95 and aa106-125. The motifs with the primary modification site Q109 are the only type of glutamine motifs present among all the antibody, B and T cell epitopes. Consistent with the roles of transglutaminase in the generation of autoimmune neoantigens[34-36, 63], these findings suggest a previously overlooked essential step mediated by transglutaminase for the determination of immunodominant epitopes in the endogenous antigen processing and presentation pathways. Given the presence of glutamine and lysine residues, an involvement of transglutaminase crosslinking will fit into the established models of immunodominance including hen-egg lysozyme (HEL) aa46-61[64, 65]. Immunodominance refers to the preference of immunological

responses for only a small subset of the peptides generated in the processing of a certain antigen protein. Origination of antigen peptides requires ubiquitin labelling of substrate proteins for their transportation to the follow-up proteolytic digestion in either proteasome or lysosome. Thus, the epitope determination within an antigen peptide would first require its proteasomal[66, 67] or lysosomal[68] survival in antigen processing. Transglutaminase crosslinking is known to render its substrates proteolytic resistance[1]. And the glutamine-lysine isopeptide modification introduced by the crosslinking enzyme is believed to prevent ubiquitination of substrate proteins via depleting lysine availability[60]. These two machineries together would ensure the modified peptides to emerge as the candidates of antigenic determinants. More importantly, the glutamine-lysine isopeptide modification over the peptides could serve as a crucial tag of determinants for their recognition by MHC and/or TCR-pMHC complex in antigen presentation. To induce a decent immune response, the determinant peptides need to bind the MHC at the greater than threshold affinity[69]. The side chain with isopeptide bond might provide better anchorage through unconventional chemical interaction with MHC residues at the interface[70]. Interestingly, the length of glutamine motifs with the highest transglutaminase reactivity in alpha-synuclein on the microchip matches the size of class I MHC binding peptides (8-10mer). However, the identification of MHC bound peptides without glutamine and lysine residues[71] imply that the dynamic interaction and positioning of the

isopeptide side chain in the TCR-pMHC complex could be the real rate-limiting step in the immunodominant determination. Based on this rationale, the TG2 modification motif determined by our assay platform should directly correspond to the substrate's epitope sequence recognized by the immunity, and our platform could be a general system applying to most epitope mapping efforts.

Amyotrophic lateral sclerosis is associated with SOD1 mutations that may greatly alter the protein's reactivity with transglutaminase and thereby contribute to the protein aggregation. With the synthesized mutant peptides on chip, our platform is also able to determine the mutation-specific epitopes through their unique TG2 modification barcodes for immunotherapeutic targeting. To a greater extent, by measuring the mutation-related crosslinking signatures our synthetic assay platform could map the disease-specific epitopes for the disorders frequently associated with genetic mutations like cancer.

Complementing the immunotherapies designed to clear the crosslinked aggregates or plaques, site-specific interfering peptides are competitive inhibitors that preventively block TG2 modification at certain sites of proteins of interest. Using an *onchip* amino-acid scanning method, we obtained 14 variants of the classic TG2 modification motif QQIV in ECM protein fibronectin with significantly higher transglutaminase reactivity. Since fibronectin crosslinking is associated with atherosclerotic disorders like myocardial infarction and stroke, and organ fibrosis, the variants could be used as the lead candidates for the development of site-specific interfering peptides treating

these complications. With the conjugation of proper fluorescent or radioactive molecules, these peptides could be further developed into diagnostic imaging agents tracing the *in situ* build-up of crosslinked plaques. We are also looking forward to applying the *onchip* amino-acid scanning method to the TG2 modification motifs in SOD1 and alpha-synuclein identified in this study. In this way an *onchip* pipeline of mapping-screening would be established for the best efficiency of synthetic and systems biology.

In this study, through scanning the residue directly following glutamine of the minimal determinant motif in the peptides QQIV and REQLYLDYNVFS, we obtained variants with significantly changed transglutaminase reactivity. To generate peptides with modified reactivity, we introduced into the fibronectin peptide QQIV additional glutamine repeats with optimized adaptor sequence as well. With either of the approaches, we manipulated TG2 modification in peptides of interest at our wills. Our results not only show the crucial role of the minimal determinant motif in the regulation of TG2 modification, but also suggest an applicable outlook for our *onchip* peptide optimization assay platform in gene editing-based protein and cell engineering. Thanks to recent advances (e.g. CRISPR/Cas9) in gene editing technology, we are able to generate *in vivo* mutant proteins with modified transglutaminase reactivity following the initial *onchip* screening in a rapid and large-scale manner. And we aspire to expand the usage of the peptide microarray-based mapping and optimization platform to studies on other posttranslational modifications including but not limited to phosphorylation[72], ubiquitination[73, 74], nitrosylation[75, 76], and many others. With the assistance of proper chemical crosslinking

reagents like disuccinimidyl tartrate (DST) and formaldehyde, our platform could be further revised to optimize protein-protein interactions by conducting large-scale protein-peptide binding assays[77]. In summary, besides epitope mapping for personalized immunotherapy, our platform of synthetic and chemical biology also provides targeting assistance to gene editing activities. We can foresee that in the near future human will be able to direct the evolution of species by on purpose generate genetically modified proteins *in vivo*.

Acknowledgement

In this draft, I would like to acknowledge Dr. Xiaolian Gao and colleagues at Department of Biology and Biochemistry, University of Houston, and Dr. Xiaochuan Zhou at LC Sciences, LLC for their assistance and inputs in peptide chip synthesis and analysis.

References:

1. Lorand L and Graham RM. Transglutaminases: crosslinking enzymes with pleiotropic functions. *Nature reviews Molecular cell biology*. 2003;4:140-56.
2. Király R, Demény M and Fésüs L. Protein transamidation by transglutaminase 2 in cells: a disputed Ca2+-dependent action of a multifunctional protein. *The FEBS journal*. 2011;278:4717-39.
3. Gundemir S, Colak G, Tucholski J and Johnson GV. Transglutaminase 2: a molecular Swiss army knife. *Biochimica et biophysica acta*. 2012;1823:406-19.
4. Eckert RL, Kaartinen MT, Nurminskaya M, Belkin AM, Colak G, Johnson GV and Mehta K. Transglutaminase regulation of cell function. *Physiological reviews*. 2014;94:383-417.
5. Greenberg CS, Birckbichler PJ and Rice RH. Transglutaminases: multifunctional cross-linking enzymes that stabilize tissues. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 1991;5:3071-7.
6. Lesort M, Tucholski J, Miller ML and Johnson GV. Tissue transglutaminase: a possible role in neurodegenerative diseases. *Progress in neurobiology*. 2000;61:439-63.
7. Muma NA. Transglutaminase is linked to neurodegenerative diseases. *Journal of neuropathology*

and experimental neurology. 2007;66:258-63.
8. Jeitner TM, Muma NA, Battaile KP and Cooper AJ. Transglutaminase activation in neurodegenerative diseases. *Future neurology*. 2009;4:449-467.
9. Grosso H and Mouradian MM. Transglutaminase 2: biology, relevance to neurodegenerative diseases and therapeutic implications. *Pharmacology & therapeutics*. 2012;133:392-410.
10. Sane DC, Kontos JL and Greenberg CS. Roles of transglutaminases in cardiac and vascular diseases. *Frontiers in bioscience : a journal and virtual library*. 2007;12:2530-45.
11. Bakker EN, Pistea A and VanBavel E. Transglutaminases in vascular biology: relevance for vascular remodeling and atherosclerosis. *Journal of vascular research*. 2008;45:271-8.
12. Liu C, Kellems RE and Xia Y. Inflammation, Autoimmunity, and Hypertension: The Essential Role of Tissue Transglutaminase. *American journal of hypertension*. 2017;30:756-764.
13. Shinde AV and Frangogiannis NG. Tissue transglutaminase in the pathogenesis of heart failure. *Cell death and differentiation*. 2018;25:453-456.
14. Sollid LM and Jabri B. Celiac disease and transglutaminase 2: a model for posttranslational modification of antigens and HLA association in the pathogenesis of autoimmune disorders. *Current opinion in immunology*. 2011;23:732-8.
15. Kárpáti S, Sárdy M, Németh K, Mayer B, Smyth N, Paulsson M and Traupe H. Transglutaminases in autoimmune and inherited skin diseases: The phenomena of epitope spreading and functional compensation. *Experimental dermatology*. 2018;27:807-814.
16. Selkoe DJ, Abraham C and Ihara Y. Brain transglutaminase: in vitro crosslinking of human neurofilament proteins into insoluble polymers. *Proceedings of the National Academy of Sciences of the United States of America*. 1982;79:6070-4.
17. Dudek SM and Johnson GV. Transglutaminase facilitates the formation of polymers of the beta-amyloid peptide. *Brain research*. 1994;651:129-33.
18. Kim SY, Grant P, Lee JH, Pant HC and Steinert PM. Differential expression of multiple transglutaminases in human brain. Increased expression and cross-linking by transglutaminases 1 and 2 in Alzheimer's disease. *The Journal of biological chemistry*. 1999;274:30715-21.
19. Norlund MA, Lee JM, Zainelli GM and Muma NA. Elevated transglutaminase-induced bonds in PHF tau in Alzheimer's disease. *Brain research*. 1999;851:154-63.
20. Balin BJ, Loewy AG and Appelt DM. Analysis of transglutaminase-catalyzed isopeptide bonds in paired helical filaments and neurofibrillary tangles from Alzheimer's disease. *Methods in enzymology*. 1999;309:172-86.
21. Rasmussen LK, Sørensen ES, Petersen TE, Gliemann J and Jensen PH. Identification of glutamine and lysine residues in Alzheimer amyloid beta A4 peptide responsible for transglutaminase-catalysed homopolymerization and cross-linking to alpha 2M receptor. *FEBS letters*. 1994;338:161-6.
22. Junn E, Ronchetti RD, Quezado MM, Kim SY and Mouradian MM. Tissue transglutaminase-induced aggregation of alpha-synuclein: Implications for Lewy body formation in Parkinson's disease and dementia with Lewy bodies. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100:2047-52.
23. Andringa G, Lam KY, Chegary M, Wang X, Chase TN and Bennett MC. Tissue transglutaminase catalyzes the formation of alpha-synuclein crosslinks in Parkinson's disease. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2004;18:932-4.
24. Kahlem P, Green H and Djian P. Transglutaminase action imitates Huntington's disease: selective polymerization of Huntingtin containing expanded polyglutamine. *Molecular cell*. 1998;1:595-601.

25. Karpuj MV, Garren H, Slunt H, Price DL, Gusella J, Becher MW and Steinman L. Transglutaminase aggregates huntingtin into nonamyloidogenic polymers, and its enzymatic activity increases in Huntington's disease brain nuclei. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96:7388-93.
26. Laki K, Benkö A and Farrell J. Clot stabilization and atherosclerosis. *Annals of the New York Academy of Sciences*. 1972;202:235-9.
27. Shainoff JR and Page IH. Deposition of modified fibrinogen within the aortic intima. *Atherosclerosis*. 1972;16:287-305.
28. Bowness JM, Folk JE and Timpl R. Identification of a substrate site for liver transglutaminase on the aminopropeptide of type III collagen. *The Journal of biological chemistry*. 1987;262:1022-4.
29. Achyuthan KE, Mary A and Greenberg CS. The binding sites on fibrin(ogen) for guinea pig liver transglutaminase are similar to those of blood coagulation factor XIII. Characterization of the binding of liver transglutaminase to fibrin. *The Journal of biological chemistry*. 1988;263:14296-301.
30. Turner PM and Lorand L. Complexation of fibronectin with tissue transglutaminase. *Biochemistry*. 1989;28:628-35.
31. Akimov SS, Krylov D, Fleischman LF and Belkin AM. Tissue transglutaminase is an integrin-binding adhesion coreceptor for fibronectin. *The Journal of cell biology*. 2000;148:825-38.
32. AbdAlla S, Lother H, Langer A, el Faramawy Y and Quitterer U. Factor XIIIA transglutaminase crosslinks AT1 receptor dimers of monocytes at the onset of atherosclerosis. *Cell*. 2004;119:343-54.
33. Liu C, Luo R, Wang W, Peng Z, Johnson GVW, Kellems RE and Xia Y. Tissue Transglutaminase-Mediated AT1 Receptor Sensitization Underlies Pro-inflammatory Cytokine LIGHT-Induced Hypertension. *American journal of hypertension*. 2019;32:476-485.
34. Bruce SE, Bjarnason I and Peters TJ. Human jejunal transglutaminase: demonstration of activity, enzyme kinetics and substrate specificity with special relation to gliadin and coeliac disease. *Clinical science (London, England : 1979)*. 1985;68:573-9.
35. Dieterich W, Ehnis T, Bauer M, Donner P, Volta U, Riecken EO and Schuppan D. Identification of tissue transglutaminase as the autoantigen of celiac disease. *Nature medicine*. 1997;3:797-801.
36. Molberg O, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, Fugger L, Scott H, Norén O, Roepstorff P, Lundin KE, Sjöström H and Sollid LM. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nature medicine*. 1998;4:713-7.
37. Liu C, Luo R, Elliott SE, Wang W, Parchim NF, Iriyama T, Daugherty PS, Blackwell SC, Sibai BM, Kellems RE and Xia Y. Elevated Transglutaminase Activity Triggers Angiotensin Receptor Activating Autoantibody Production and Pathophysiology of Preeclampsia. *Journal of the American Heart Association*. 2015;4.
38. Gentile V and Cooper AJ. Transglutaminases - possible drug targets in human diseases. *Current drug targets CNS and neurological disorders*. 2004;3:99-104.
39. Wodzinska JM. Transglutaminases as targets for pharmacological inhibition. *Mini reviews in medicinal chemistry*. 2005;5:279-92.
40. Siegel M and Khosla C. Transglutaminase 2 inhibitors and their therapeutic role in disease states. *Pharmacology & therapeutics*. 2007;115:232-45.
41. Jeitner TM, Pinto JT and Cooper AJL. Cystamine and cysteamine as inhibitors of transglutaminase activity in vivo. *Bioscience reports*. 2018;38.
42. Zhuang R and Khosla C. Substrates, inhibitors, and probes of mammalian transglutaminase 2.

Analytical biochemistry. 2020;591:113560.
43. Sugimura Y, Hosono M, Wada F, Yoshimura T, Maki M and Hitomi K. Screening for the preferred substrate sequence of transglutaminase using a phage-displayed peptide library: identification of peptide substrates for TGASE 2 and Factor XIIIA. *The Journal of biological chemistry*. 2006;281:17699-706.
44. Keresztessy Z, Csosz E, Hársfalvi J, Csomós K, Gray J, Lightowlers RN, Lakey JH, Balajthy Z and Fésüs L. Phage display selection of efficient glutamine-donor substrate peptides for transglutaminase 2. *Protein science : a publication of the Protein Society*. 2006;15:2466-80.
45. Hitomi K, Kitamura M and Sugimura Y. Preferred substrate sequences for transglutaminase 2: screening using a phage-displayed peptide library. *Amino acids*. 2009;36:619-24.
46. Nemes Z, Devreese B, Steinert PM, Van Beeumen J and Fésüs L. Cross-linking of ubiquitin, HSP27, parkin, and alpha-synuclein by gamma-glutamyl-epsilon-lysine bonds in Alzheimer's neurofibrillary tangles. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2004;18:1135-7.
47. Konno T, Morii T, Hirata A, Sato S, Oiki S and Ikura K. Covalent blocking of fibril formation and aggregation of intracellular amyloidgenic proteins by transglutaminase-catalyzed intramolecular cross-linking. *Biochemistry*. 2005;44:2072-9.
48. Nemes Z, Petrovski G, Aerts M, Sergeant K, Devreese B and Fésüs L. Transglutaminase-mediated intramolecular cross-linking of membrane-bound alpha-synuclein promotes amyloid formation in Lewy bodies. *The Journal of biological chemistry*. 2009;284:27252-64.
49. Tatsukawa H, Tani Y, Otsu R, Nakagawa H and Hitomi K. Global identification and analysis of isozyme-specific possible substrates crosslinked by transglutaminases using substrate peptides in mouse liver fibrosis. *Scientific reports*. 2017;7:45049.
50. Lee JH, Song C, Kim DH, Park IH, Lee SG, Lee YS and Kim BG. Glutamine (Q)-peptide screening for transglutaminase reaction using mRNA display. *Biotechnology and bioengineering*. 2013;110:353-62.
51. Malešević M, Migge A, Hertel TC and Pietzsch M. A fluorescence-based array screen for transglutaminase substrates. *Chembiochem : a European journal of chemical biology*. 2015;16:1169-74.
52. Csosz E, Bagossi P, Nagy Z, Dosztanyi Z, Simon I and Fesus L. Substrate preference of transglutaminase 2 revealed by logistic regression analysis and intrinsic disorder examination. *Journal of molecular biology*. 2008;383:390-402.
53. Csosz E, Meskó B and Fésüs L. Transdab wiki: the interactive transglutaminase substrate database on web 2.0 surface. *Amino acids*. 2009;36:615-7.
54. Tagami U, Shimba N, Nakamura M, Yokoyama K, Suzuki E and Hirokawa T. Substrate specificity of microbial transglutaminase as revealed by three-dimensional docking simulation and mutagenesis. *Protein engineering, design & selection : PEDS*. 2009;22:747-52.
55. Lorand L, Urayama T, De Kiewiet JW and Nossel HL. Diagnostic and genetic studies on fibrin-stabilizing factor with a new assay based on amine incorporation. *The Journal of clinical investigation*. 1969;48:1054-64.
56. Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E and Gao X. Individually addressable parallel peptide synthesis on microchips. *Nature biotechnology*. 2002;20:922-6.
57. Houghten RA, Pinilla C, Blondelle SE, Appel JR, Dooley CT and Cuervo JH. Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature*. 1991;354:84-6.

58. Nazif T and Bogyo M. Global analysis of proteasomal substrate specificity using positional-scanning libraries of covalent inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98:2967-72.
59. Parameswaran KN, Velasco PT, Wilson J and Lorand L. Labeling of epsilon-lysine crosslinking sites in proteins with peptide substrates of factor XIIIa and transglutaminase. *Proceedings of the National Academy of Sciences of the United States of America*. 1990;87:8472-5.
60. Liu C, Wang W, Parchim N, Irani RA, Blackwell SC, Sibai B, Jin J, Kellems RE and Xia Y. Tissue transglutaminase contributes to the pathogenesis of preeclampsia and stabilizes placental angiotensin receptor type 1 by ubiquitination-preventing isopeptide modification. *Hypertension (Dallas, Tex : 1979)*. 2014;63:353-61.
61. Masliah E, Rockenstein E, Adame A, Alford M, Crews L, Hashimoto M, Seubert P, Lee M, Goldstein J, Chilcote T, Games D and Schenk D. Effects of alpha-synuclein immunization in a mouse model of Parkinson's disease. *Neuron*. 2005;46:857-68.
62. Davtyan H, Zagorski K, Petrushina I, Kazarian K, Goldberg NRS, Petrosyan J, Blurton-Jones M, Masliah E, Cribbs DH, Agadjanyan MG and Ghochikyan A. MultiTEP platform-based DNA vaccines for alpha-synucleinopathies: preclinical evaluation of immunogenicity and therapeutic potency. *Neurobiology of aging*. 2017;59:156-170.
63. Ellis HJ, Pollock EL, Engel W, Fraser JS, Rosen-Bronson S, Wieser H and Ciclitira PJ. Investigation of the putative immunodominant T cell epitopes in coeliac disease. *Gut*. 2003;52:212-7.
64. Allen PM, Strydom DJ and Unanue ER. Processing of lysozyme by macrophages: identification of the determinant recognized by two T-cell hybridomas. *Proceedings of the National Academy of Sciences of the United States of America*. 1984;81:2489-93.
65. Babbitt BP, Allen PM, Matsueda G, Haber E and Unanue ER. Binding of immunogenic peptides to Ia histocompatibility molecules. *Nature*. 1985;317:359-61.
66. Goldberg AL and Rock KL. Proteolysis, proteasomes and antigen presentation. *Nature*. 1992;357:375-9.
67. Michalek MT, Grant EP, Gramm C, Goldberg AL and Rock KL. A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. *Nature*. 1993;363:552-4.
68. Trombetta ES, Ebersold M, Garrett W, Pypaert M and Mellman I. Activation of lysosomal function during dendritic cell maturation. *Science (New York, NY)*. 2003;299:1400-3.
69. Yewdell JW and Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annual review of immunology*. 1999;17:51-88.
70. Fremont DH, Monnaie D, Nelson CA, Hendrickson WA and Unanue ER. Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. *Immunity*. 1998;8:305-17.
71. Rudensky Y, Preston-Hurlburt P, Hong SC, Barlow A and Janeway CA, Jr. Sequence analysis of peptides bound to MHC class II molecules. *Nature*. 1991;353:622-7.
72. Ptacek J, Devgan G, Michaud G, Zhu H, Zhu X, Fasolo J, Guo H, Jona G, Breitkreutz A, Sopko R, McCartney RR, Schmidt MC, Rachidi N, Lee SJ, Mah AS, Meng L, Stark MJ, Stern DF, De Virgilio C, Tyers M, Andrews B, Gerstein M, Schweitzer B, Predki PF and Snyder M. Global analysis of protein phosphorylation in yeast. *Nature*. 2005;438:679-84.
73. Merbl Y and Kirschner MW. Large-scale detection of ubiquitination substrates using cell extracts and protein microarrays. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106:2543-8.
74. Merbl Y, Refour P, Patel H, Springer M and Kirschner MW. Profiling of ubiquitin-like modifications

reveals features of mitotic control. *Cell*. 2013;152:1160-72.
75. Foster MW, Forrester MT and Stamler JS. A protein microarray-based analysis of S-nitrosylation. *Proceedings of the National Academy of Sciences of the United States of America*. 2009;106:18948-53.
76. Lee YI, Giovinazzo D, Kang HC, Lee Y, Jeong JS, Doulias PT, Xie Z, Hu J, Ghasemi M, Ischiropoulos H, Qian J, Zhu H, Blackshaw S, Dawson VL and Dawson TM. Protein microarray characterization of the S-nitrosoproteome. *Molecular & cellular proteomics : MCP*. 2014;13:63-72.
77. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M and Snyder M. Global analysis of protein activities using proteome chips. *Science (New York, NY)*. 2001;293:2101-5.

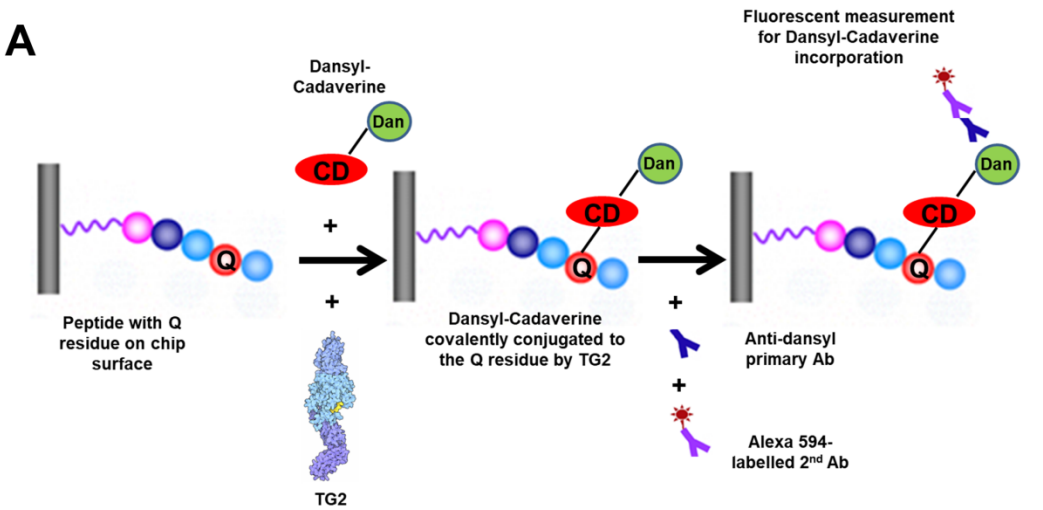
Figure Legend

Figure 1. Peptide microarray-based *in vitro* TG2 assay platform for rapid and high-throughput identification of modification Q sites among proteins of interest. A. Assay flow chart for the Q mapping platform (Q: glutamine; CD: cadaverine; Dan: dansyl; Ab: antibody; TG2: tissue transglutaminase). **B.** Peptide layout strategy and naming rule on the microchip using the first glutamine (Q24) motif in human alpha-synuclein as the example. Each peptide where position of Q is defined as 0 and the one directly before or after it as -1 or 1 is named as *protein name*_Q *residue# in the entire protein_N terminal residue# in the peptide~C terminal residue# in the peptide*.

Figure 2. Identification of the glutamine-bearing TG2 modification sites in SOD1 and alpha-synuclein with the high-throughput *in vitro* assay platform. A. Representative image of the protein microarray-based *in vitro* TG2 assay with the Q peptides in SOD1 showing that the first Q (Q16) peptides (red box) have a generally stronger reaction signal (red box) than the other two (Q23:yellow box; Q154:green box). Q16 residue in SOD1 (**B**) and Q109 in alpha-synuclein (**C**) are identified as the primary TG2 modification sites ($p<0.05$ vs other residues or mutant controls in **B** and **C**; n=triplicates; data=mean ± SEM). The top 3 peptide hits of each Q residue in the TG2 reaction are plotted together with their Q to S mutants. **D.** Alignment of the top 10 hits of Q16 peptides in SOD1 and Q109 peptides in alpha-synuclein indicate their minimal determinant motifs.

Figure 3. *Onchip* amino-acid scanning generates peptides with significantly changed reactivity with TG2. A. Pan-amino acid swapping at the I site of the peptide sequence QQIV generates 14 variants (red bars) with significantly higher TG2 reactivity and 1 lower (dark green) (*P<0.05 versus QQIV; n=triplicates; data=mean ± SEM). **B.** Amino acid scanning at the L residue directly after the Q site of the TG2 substrate peptide REQLYLDYNVFS also obtained variants with significantly changed TG2 reactivity (*P<0.05 versus REQLYLDYNVFS; n=triplicates; data=mean ± SEM).

Figure 4. Addition of glutamine repeat elevates TG2 reactivity of the fibronectin peptide QQIV. A. Addition of glutamine repeat at the N terminal of the fibronectin peptide QQIV results in significant increase in dansyl-cadaverine incorporation as measured by fluorescent intensity on the chip (**P<0.01 versus QQIV; X=any of 20 amino acids; n=triplicates; data=mean ± SEM). **B.** Majority (297 out of total 400) of the QQXXQQIV variants show a higher level of TG2 reactivity as measured by fluorescent intensity on the chip (the fluorescent intensity of QQIV is ~5000 as indicated). **C.** Among the top 40 QQXXQQIV peptides with the highest fluorescent intensity, the peptides with a P at the third residue or I fourth appear most.



B

AEKTK**Q**GVAEAA

-5 -4 -3 -2 -1 0 1 2 3 4 5 6

$\mathbf{Q}_{24}$GVA:α-synuclein_Q24_-0~3
K$\mathbf{Q}_{24}$GV:α-synuclein_Q24_-1~2
TK$\mathbf{Q}_{24}$G:α-synuclein_Q24_-2~1
KTK$\mathbf{Q}_{24}$:α-synuclein_Q24_-3~0

$\mathbf{Q}_{24}$GVAE:α-synuclein_Q24_-0~4
K$\mathbf{Q}_{24}$GVA:α-synuclein_Q24_-1~3
TK$\mathbf{Q}_{24}$GV:α-synuclein_Q24_-2~2
KTK$\mathbf{Q}_{24}$G:α-synuclein_Q24_-3~1
EKTK$\mathbf{Q}_{24}$:α-synuclein_Q24_-4~0

Figure 1

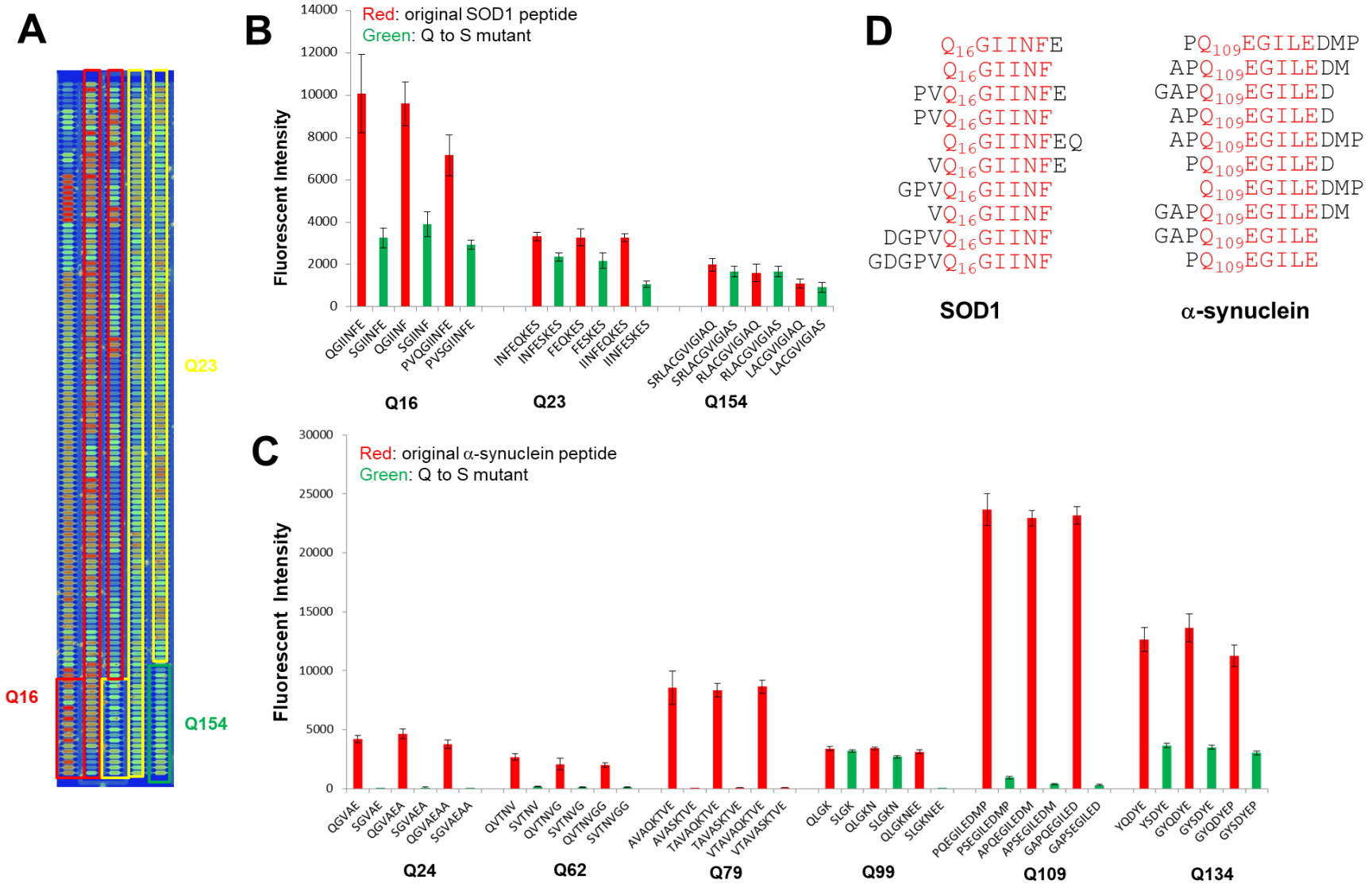

A
Q23
Q16
Q154
B
Red: original SOD1 peptide
Green: Q to S mutant
Fluorescent Intensity
14000
12000
10000
8000
6000
4000
2000
0
QGIINFE
SGIINFE
QGIINF
SGIINF
PVQGIINFE
PVSGIINFE
INFEQKES
INFESKES
FEQKES
FESKES
IINFEQKES
IINFESKES
SRLACGVIGIAQ
SRLACGVIGIAS
RLACGVIGIAQ
RLACGVIGIAS
LACGVIGIAQ
LACGVIGIAS
Q16
Q23
Q154
C
Red: original α-synuclein peptide
Green: Q to S mutant
Fluorescent Intensity
30000
25000
20000
15000
10000
5000
0
QGVAE
SGVAE
QGVAEA
SGVAEA
QGVAEAA
SGVAEAA
QVTNV
SVTNV
QVTNVG
SVTNVG
QVTNVGG
SVTNVGG
AVAQKTVE
AVASKTVE
TAVAQKTVE
TAVASKTVE
VTAVAQKTVE
VTAVASKTVE
QLGK
SLGK
QLGKN
SLGKN
QLGKNEE
SLGKNEE
PQEGILEDMP
PSEGILEDMP
APQEGILEDM
APSEGILEDM
GAPQEGILED
GAPSEGILED
YQDYE
YSDYE
GYQDYE
GYSDYE
GYQDYEP
GYSDYEP
Q24
Q62
Q79
Q99
Q109
Q134
D
Q16GIINFE
Q16GIINF
PVQ16GIINFE
PVQ16GIINF
Q16GIINFEQ
VQ16GIINFE
GPVQ16GIINF
VQ16GIINF
DGPVQ16GIINF
GDGPVQ16GIINF
SOD1
PQ109EGILEDMP
APQ109EGILEDM
GAPQ109EGILED
APQ109EGILED
APQ109EGILEDMP
PQ109EGILED
Q109EGILEDMP
GAPQ109EGILEDM
GAPQ109EGILE
PQ109EGILE
α-synuclein

Figure 2

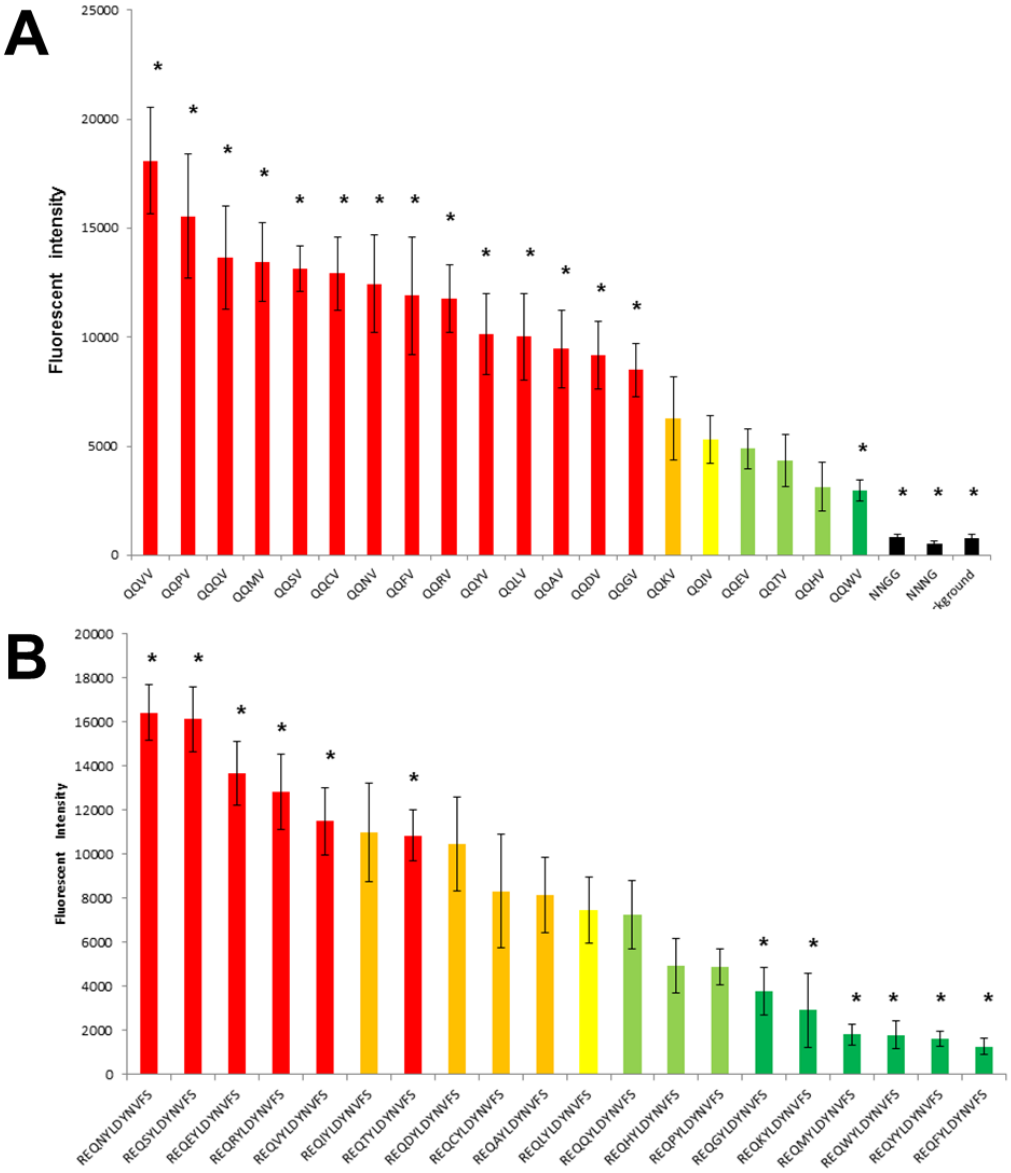

A
Fluorescent intensity
25000
20000
15000
10000
5000
0
QQVV
QQPV
QQQV
QQMV
QQSV
QQCV
QQNV
QQFV
QQRV
QQYV
QQLV
QQAV
QQDV
QQGV
QQKV
QQIV
QQEV
QQTV
QQHV
QQWV
NNGG
NNNG
B
Fluorescent Intensity
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
REQNYLDYNVFS
REQSYLDYNVFS
REQEYLDYNVFS
REQRYLDYNVFS
REQVYLDYNVFS
REQIYLDYNVFS
REQTYLDYNVFS
REQDYLDYNVFS
REQCYLDYNVFS
REQAYLDYNVFS
REQLYLDYNVFS
REQQYLDYNVFS
REQHYLDYNVFS
REQPYLDYNVFS
REQGYLDYNVFS
REQKYLDYNVFS
REQMYLDYNVFS
REQWYLDYNVFS
REQYYLDYNVFS
REQFYLDYNVFS

Figure 3

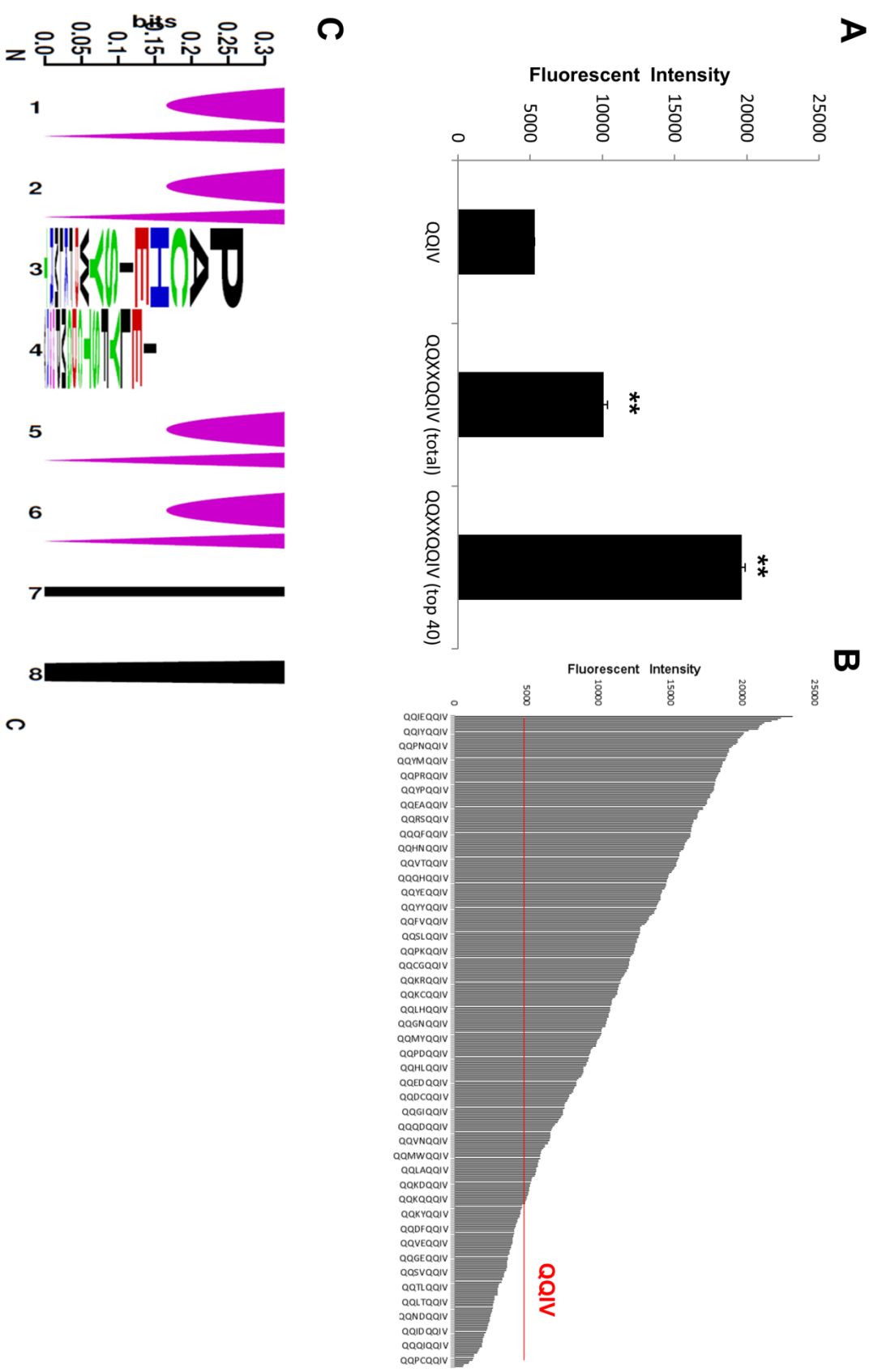

A
Fluorescent Intensity
25000
20000
15000
10000
5000
0
QQIV
QQXXQQIV (total)
QQXXQQIV (top 40)
**
**
B
Fluorescent Intensity
QQIV
C
bits
N
1
2
3
4
5
6
7
8

Figure 4