(12) STANDARD PATENT (11) Application No. AU 2019257105 B2
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Protein macrocyclization

(51) International Patent Classification(s)
C12N 9/96 (2006.01) ***C12N 9/64*** (2006.01)
C12N 9/52 (2006.01)

(21) Application No: **2019257105** (22) Date of Filing: **2019.04.18**

(87) WIPO No: **WO19/203645**

(30) Priority Data

(31) Number	(32) Date	(33) Country
18168298.0	**2018.04.19**	**EP**

(43) Publication Date: **2019.10.24**
(44) Accepted Journal Date: **2024.11.28**

(71) Applicant(s)
Stichting VU

(72) Inventor(s)
GROSSMANN, Tom Norbert;PELAY GIMENO, Marta;HENNIG, Sven;NEUBACHER, Saskia Antonie

(74) Agent / Attorney
IP GATEWAY PATENT & TRADE MARK ATTORNEYS PTY LTD, PO Box 1321, SPRINGWOOD, QLD, 4127, AU

(56) Related Art
NASIR KATO BASHIRUDDIN ET AL: "Synthesis of fused tricyclic peptides using a reprogrammed translation system and chemical modification", BIOORGANIC CHEMISTRY., vol. 61, 1 Aug 2015, pages 45-50, DOI: 10.1016/j.bioorg.2015.06.002
PHILIPPE DIDERICH ET AL: "Directed Evolution of Bicyclic Peptides for Therapeutic Application", CHIMIA INTERNATIONAL JOURNAL FOR CHEMISTRY, vol. 67, no. 12, 18 December 2013, pages 910 - 915, DOI: 10.2533/chimia.2013.910
ALESSANDRO ANGELINI ET AL: "Chemical Macrocyclization of Peptides Fused to Antibody Fc Fragments", BIOCONJUGATE CHEMISTRY, vol. 23, no. 9, 19 September 2012, pages 1856 - 1863, DOI: 10.1021/bc300184m
SHIYU CHEN ET AL: "Structurally Diverse Cyclisation Linkers Impose Different Backbone Conformations in Bicyclic Peptides", CHEMBIOCHEM, vol. 13, no. 7, 7 May 2012, pages 1032-1038, DOI: 10.1002/cbic.201200049

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
24 October 2019 (24.10.2019)

WIPO | PCT

(10) **International Publication Number**
WO 2019/203645 A1

(51) **International Patent Classification:**
C12N 9/96 (2006.01) *C12N 9/64* (2006.01)
C12N 9/52 (2006.01)

(21) **International Application Number:**
PCT/NL2019/050229

(22) **International Filing Date:**
18 April 2019 (18.04.2019)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
18168298.0 19 April 2018 (19.04.2018) EP

(71) **Applicant: STICHTING VU** [NL/NL]; De Boelelaan 1105, 1081 HV Amsterdam (NL).

(72) **Inventors: GROSSMANN, Tom Norbert**; c/o De Boelelaan 1105, 1081 HV Amsterdam (NL). **PELAY GIMENO, Marta**; c/o De Boelelaan 1105, 1081 HV Amsterdam (NL). **HENNIG, Sven**; c/o De Boelelaan 1105, 1081 HV Amsterdam (NL). **NEUBACHER, Saskia Antonie**; c/o De Boelelaan 1105, 1081 HV Amsterdam (NL).

(74) **Agent: JANSEN, C.M.**; P.O. Box 87930, 2508 DH Den Haag (NL).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) **Title:** PROTEIN MACROCYCLIZATION

(57) **Abstract:** The present invention relates to methods and cross-linkers for the macrocyclization of proteins. The invention is useful for increasing the stability of a protein.

Title: Protein macrocyclization

FIELD OF THE INVENTION

The present invention relates to methods and trivalent thiol-reactive cross-linkers for the macrocyclization of proteins. The invention is useful for increasing the stability of a protein.

DEFINITION

In this specification, the term “comprising” is intended to denote the inclusion of a stated integer or integers, but not necessarily the exclusion of any other integer, depending on the context in which that the term is used. This applies to variants of that term such as “comprising” or “comprises”.

BACKGROUND OF THE INVENTION

The reference to prior art in this specification is not and should not be taken as an acknowledgment or any form of suggestion that the referenced prior art forms part of the common general knowledge in Australia or in any other country.

Enzymes are an essential component of most biotechnological and biomedical processes[1,2] but their scope of application is hampered by a limited stability under often desired harsh conditions (e.g. elevated temperature or presence of denaturants). Consequently, the stabilization of protein structures is a central aspect in the development of suitable enzymes. The complexity of interactions in protein tertiary structures and the sensitivity of enzymatic activity on sequence alterations render enzyme stabilization very challenging. A minimal invasive strategy involves the use of covalent protein modifications (e.g. pegylation or glycosylation) being mainly applied to increase biostability for therapeutic applications.[3,4] Alternatively, enzyme stabilization can be achieved via alterations in the protein sequence applying directed evaluation, consensus-based mutagenesis or computational approaches[5,6,7,8] which can be complimented by the introduction of non-proteinogenic amino acids.[9] These approaches aim for improved protein core interactions, structure rigidification, and/or surface charge distribution and often require multiple rounds of optimization to achieve relevant stabilization effects.

There exists a need for new methods to increase the stability of proteins, in particular enzymes.

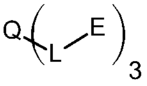
SUMMARY OF THE INVENTION

In one aspect the disclosure provides a method for increasing the stability of a protein, wherein the protein comprises at least 70 amino acids, said method comprising:

- a) providing a protein comprising three cysteine residues and
- b) contacting said protein with a trivalent thiol-reactive cross-linker such that the linker forms covalent bonds with each of the three cysteine residues.

Preferably, step a) comprises modifying a protein to introduce one or more of the three cysteine residues. Preferably, said protein comprises at least a fourth cysteine residue and wherein said method does not result in the formation of a covalent bond between the fourth cysteine and the cross-linker. Preferably, the protein is an enzyme and the fourth cysteine is part of the enzymatic active site. Preferably, the cross-linker has C_3 symmetry.

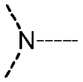
Preferably, the cross-linker has formula (I):

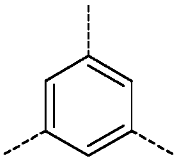
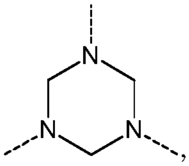


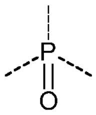
(I)

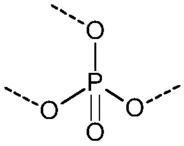
wherein

Q is a core structure selected from the group consisting of

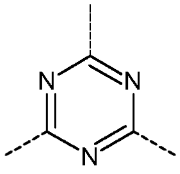


,

,


,


,


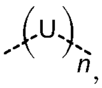
, and

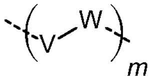


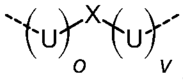
,

each dashed line in Q indicating a site where Q is bound to L,

each L is a linker independently selected from the group consisting of



,


, and


,

wherein

each U is independently selected from CH_2 and CF_2,

V is CH_2

W is CF_2

X is NR, NH or O,

wherein R is a fluorophore or affinity handle,

n is an integer in the range of 2-8,

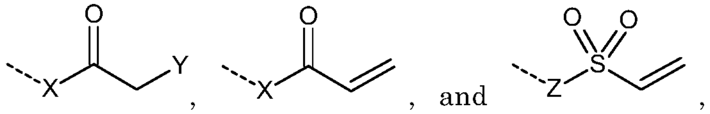
m is an integer in the range of 1-4

o is 2 or 3, and

v is 2 or 3,

each dashed line in L indicating a site where L is bound to Q or E,

each E is an electrophile independently selected from the group consisting of



wherein

each X is independently selected from NH and O,

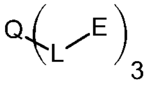
Y is selected from F, Cl, Br, Tos ($O-SO_2-C_6H_4-CH_3$), and Mes ($O-SO_2-CH_3$),

Z is CH_2, $NH-C(O)-CH_2$, or $O-C(O)-CH_2$

each dashed line in E indicating a site where E is bound to L.

The disclosure further provides a stabilized protein obtainable by a method as disclosed herein. Preferably the protein is a Sortase A polypeptide or a KIX domain polypeptide.

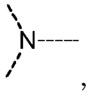
The disclosure further provides a trivalent thiol-reactive cross-linker having formula II:



(II)

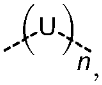
wherein

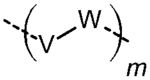
Q is

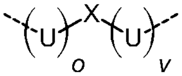


each dashed line in Q indicating a site where Q is bound to L,

each L is a linker independently selected from the group consisting of



,


, and


,

wherein

each U is independently selected from CH_2 and CF_2,

V is CH_2

W is CF_2

X is NR, NH or O,

wherein R is a fluorophore or affinity handle, n is an integer in the range of 2-8,

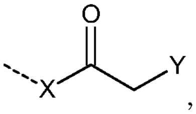
m is an integer in the range of 1-4

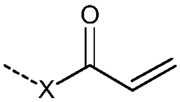
o is 2 or 3, and

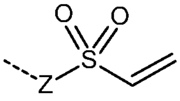
v is 2 or 3,

each dashed line in L indicating a site where L is bound to Q or E,

each E is an electrophile independently selected from the group consisting of



,


, and


,

wherein

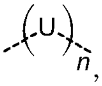
each X is independently selected from NH and O,

Y is selected from F, Cl, Br, Tos ($O-SO_2-C_6H_4-CH_3$), and Mes ($O-SO_2-CH_3$),

Z is CH_2, $NH-C(O)-CH_2$, or $O-C(O)-CH_2$

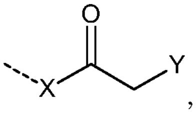
each dashed line in E indicating a site where E is bound to L.

Preferably,

L is


,

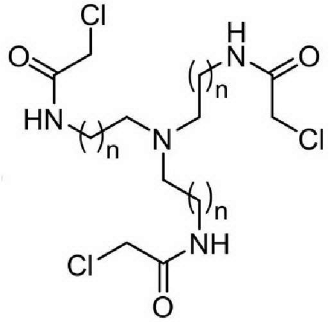
U is CH_2, preferably n is 2 or 3; and

E is


,

wherein X is NH and Y is F, Cl, or Br, preferably Cl.

Preferably the cross-linker is Formula III:



wherein n is 1 or n is 2.

The disclosure further provides the use of the cross-linkers disclosed herein for reacting with thiol groups, preferably for cross-linking three cysteine resides present in a protein.

The disclosure further provides a protein comprising at least 70 amino acids and comprising at least three cysteine residues, wherein each of the three cysteine residues is covalently bonded to a trivalent thiol-reactive cross-linker. Preferably, the trivalent thiol-reactive cross-linker is a cross-linker as disclosed herein.

The disclosure further provides a Sortase A polypeptide comprising amino acid substitutions with cysteine at positions 111, 149, and 177, with reference to amino acid position numbering of Staphylococcus aureus SrtA, preferably wherein the polypeptide has SEQ ID NO: 1.

The disclosure further provides a KIX domain polypeptide comprising amino acid substitutions with cysteine at positions 594, 599, and 646, with reference to amino acid position numbering of Figure 13b, preferably wherein the polypeptide has the sequence depicted in Figure 13c.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. a) Macrocyclization strategy towards stabilized protein tertiary structures using a modular bis- or triselectrophilic crosslink; b) Electrophiles (maleimide 1, 2-bromoacetamide 2, 2-chloroacetamide 3, acrylamide 4) considered for crosslinking of accessible cysteines.

Figure 2. a) Biselectrophiles (b1 – b6) used for the generation of cyclic enzymes; b) NMR structure of SrtA (PDB: 1ija) with positions of cysteine variations high-lighted. Cysteine pairs (same color) and their positions are shown; c) Heat map representation

of Tm-values for linear and crosslinked SrtA variants (75 µM); d) Mechanism for SrtA-mediated transpeptidation reactions (recognition motif: LPETG); e) Heat map representation of enzymatic activity (vr, relative to wildtype SrtA) of linear and crosslinked SrtA variants at 65 °C (10 µM enzyme, 10 µM fluorescent probe) (buffer for 2c and 2e: 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP with 0.01 % Tween for 3e).

Figure 3. a) NMR structure of SrtA (PDB: 1ija) with positions of cysteine variations in S7 highlighted; b) Chemical structure of triselectrophile t1 and Coomassie-stained SDS-PAGE gel showing protein bands after incubation t1 (50 µM S7, 1 mM t1, 50 mM HEPES, pH 8.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP); c) Melting curves of SrtA, S4-b3 and S7-t1 including Tm-values; d) Fluorescent readout of probe cleavage upon enzyme activity at 65 °C (10 µM enzyme, 10 µM fluorescent probe) (buffer for 3c and 3d: 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP with 0.01 % Tween for d).

Figure 4. a) HPLC chromatograms (440 nm) of transpeptidation reaction (12 h at 65 °C) with fluorescence probe (●) in absence of enzyme (light grey), with SrtA (dark grey) or with S7-t1 (red). Product formation (▲, ■) was only observed in presence of S7-t1 (50 µM enzyme, 10 µM probe, 2.5 mM GGG); b) Temperature dependence of enzymatic activity (vr, relative to SrtA at 37 °C) (10 µM enzyme, 10 µM probe, 2.5 mM GGG, buffer: 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP, 0.01 % Tween). Values are mean of triplicate (+/- 1 σ, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, ns: not significant); c) Relative enzymatic activity (vr, relative to SrtA in absence of GdnHCl) at 37 °C for various concentrations of GdnHCl (for conditions and data processing see 3b); d) Coomassie-stained SDS-PAGE gel showing the soluble α-Syn fractions before fibril formation (A) and after re-solubilization in absence (B) and presence (C) of GdnHCl (1 M); e) Fluorescent readout (λem = 520 nm) of α Syn labeling using soluble fractions before fibril formation A (w/o GdnHCl) and after re-solubilization B (w/o GdnHCl) and C (1 M GdnHCl) with either SrtA or S7-t1.

Figure 5: Chemical structure of biselectrophilic cross-linkers and their calculated lengths using 3D ChemDraw.

Figure 6: a) Amino acid sequence of SrtA cloned, expressed and evaluated in the present study; b) List of StrA variants (highlighted in the corresponding color code

(see Figure 2b)) and measured distances between Cα atoms in the NMR structure (PDB code: 1ija).

Figure 7: a) Schematic depiction of SrtA enzymatic activity assay. Chemical structure, HPLC chromatogram (linear gradient from 30% to 70% ACN over 10 min, 210 nm) and MS spectra of fluorescence probe. b) Plot of SrtA hydrolysis activity assay at 37 °C (solid line) and 65°C (dashed line).

Figure 8: Analysis of the modification state of the four cysteines in S7 and S7-t1 by HPLC-coupled high-resolution mass spectrometry. The spectra for the following tryptic fragments are shown:

a) aa 100 – 134 (including C111), b) 138 – 151 (including C149), c) 176 – 190 (including C177 and active site C184). Free cysteine residues or C_4H_4ON (grey) modified ones indicate a free cysteine in the enzyme (S7 or S7-t1) before MS workup, while C_2H_2O (red) modified cysteines indicate a covalent modification with t1.

Figure 9: a) SEQ ID NO:2: α-Synuclein (α-Syn) sequence including a C-terminal flexible linker (bold), SrtA recognition motif (grey) and a His6-tag (underlined) for affinity purification. b) Chemical structure and HPLC/MS analysis of the fluorescent probe GGGK-FITC. Chromatogram (λ = 210 nm) and MS spectrum are shown.

Figure 10: a) Scheme of fibril formation and re-solubilization with GdnHCl (A: purified α-Syn prior to fibril formation, B: re-solubilization attempt in absence of GdnHCl - insoluble α-Syn fibrils, C: re-solubilized α-synuclein following GdnHCl (1 M) treatment); b) Coomassie-stained (I) and fluorescent readout (II) of full SDS-PAGE gel (17% SDS) shown in Figure 4d and 4e.

Figure 11: Transpeptidase activity of StrA (grey) and bicyclic S7-t1 (red) before and after thermal denaturation. Untreated proteins (StrA and S7-t1) and samples subjected to a heating/cooling cycle (StrA# and S7-t1#, heating from r.t. to 85 °C over 30 min and cooled to r.t. over 15 min) are compared. A sample without enzyme was included (-). Statistical significance was evaluated by an unpaired t test ($n = 3$, ns: not significant $p > 0.05$)..

Figure 12. a) NMR structure of KIX (PDB: 2agh) with positions of cysteine variations in K1 highlighted. The secondary structural elements have been named; b) Melting curves of KIX, K1-t1 and K1-t2 including Tm-values.

Figure 13: a) Amino acid sequence of KIX wt sub-cloned, expressed and evaluated in the present study. b) Amino acid sequence of KIX variant K1 sub-cloned, expressed

and evaluated in the present study. Mutations are highlighted in green. c) Averaged distance between the Cα atoms of underlined amino acid positions over the 20 conformers of the KIX NMR structure (PDB code: 2agh). d) Cysteine positions in S7 variant. Averaged distance between the Cα atoms of underlined amino acid positions over the 25 conformers of the SrtA NMR structure (PDB code: 1ija).

Figure 14: a) Chemical structure, HPLC chromatogram (linear gradient from 40 % to 80 % ACN over 10 min (3-13 min), λ = 210 nm) and MS spectra of MLL peptide. b) FP assay of KIX wt and crosslinked versions K1-t1 and K1-t2. The corresponding Kd values are shown.

Figure 15: a) Coomassie-stained SDS-PAGE gel showing protein bands after incubation with t2 (50 μM S7, 1 mM t2, 50 mM HEPES, pH 8.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP). b) Table with calculated and found m/z-values for S7-t2. c) MS spectra of bicyclic S7-t2. d) Melting curves of SrtA and S7-t2 including apparent Tm-values. e) Fluorescent readout of probe cleavage upon enzyme hydrolytic activity at 65 °C (10 μM enzyme, 10 μM fluorescent probe). Buffer used for experiments d and e: 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP, with 0.01% Tween 20 for e).

DETAILED DESCRIPTION OF THE DISCLOSED EMBODIMENTS

Increasing the stabilization of enzymes against denaturants and elevated temperature has also been studied by the introduction of intramolecular crosslinks. There have been reports of the installation of additional disulfide bridges or, as recently reported for non-enzymatic protein domains, the introduction of disulfide mimics[10] which are insensitive to reducing environments. In addition, the crosslinking of protein termini via lactam formation was applied[11-13] requiring a suitable spatial alignment of N- and C-terminus in the tertiary structure.

To reduce these structural prerequisites, the incorporation of non-natural electrophilic amino acids was pursued to enable crosslinking with appropriately aligned cysteine side chains.[14] However, the use of amber stop codon suppression for the introduction of these non-natural amino acids complicates protein expression. In addition, the screening of linker libraries is hampered since incorporation of these modified amino acids requires adapted tRNA synthetases which is work intensive and does not succeed for every non-natural amino acid.[15] As such, cross-linking approaches that rely on the use of non-natural amino acids suffer a number of disadvantages.

The present disclosure provides methods for producing cross-linked proteins that do not rely on the use of non-natural amino acids. The protein may have any function, e.g., cytokines, chemokines, growth factors, hormones, antibodies, receptors, and antigens, etc. In some embodiments, the protein is an enzyme. As is apparent to a skilled person, the cross-linked proteins described herein include proteins composed of more than one polypeptide or peptide chain. For example, antibodies are an exemplary protein of the disclosure and IgG antibodies are made up of four peptide chains. Receptors are also exemplary proteins of the disclosure and many proteins are made up of, e.g., hetero- or homo-dimers. In some embodiments, the cysteine residues are present in different polypeptides or peptide chains, such that the cross-links are formed between polypeptide/peptide chains. For example, the present disclosure contemplates methods for increasing the stability of a hetero-dimer receptor. As an example of one of the many embodiments contemplated herein, a hetero-dimer receptor is provided having one cysteine residue in one receptor subunit and two cysteine residues in the other receptor subunit such that when the receptor is contacted with a trivalent thiol-reactive cross-linker the linker forms covalent bonds with each of the three cysteine residues and between the dimers.

In particular, the methods improve the stability of a protein using cross-linking reagents. As is clear to a skilled person, an increase in stability refers to an increase in stability of the cross-linked protein as compared to the stability of the protein without cross-linking. The term “increased stability” as used herein, refers to -an increase in resistance to- or –a decrease in susceptibility to- denaturation. Denaturation refers to the loss of secondary or tertiary structure and the biological function, in particular enzymatic activity, of most proteins is reduced or lost when denatured. Denaturation can occur as a result of mechanical agitation, radiation, increased temperature, or by chemical denaturants. In some embodiments, improved stability refers to the presence of a higher ratio of folded to unfolded protein when cross-linked, relative to that of the protein without cross-linking. Improved stability can be determined by examining the amount of folded protein present under varying conditions, e.g., temperature, detergent, denaturing agents, and pH. In preferred embodiments, the methods improve the thermal stability and/or stability against chemical denaturants.

In some embodiments, the stability of the protein can be determined by measuring the Tm. The term “Tm” refers to the temperature at which 50% of the protein has unfolded. Typically, the higher the Tm, the more stable the protein. In some embodiments, the methods are for increasing the Tm of a protein.
In some embodiments, the stability of the protein can be determined by measuring the effects of chemical agents on the protein. Chemical denaturants are agents that can disrupt non-covalent interactions and covalent bonds within a protein. Exemplary chemical denaturants include guanidinium hydrochloride, guanadinium thiocyanate, urea, acetone, organic solvents, salts, reducing agents (e.g. dithiothreitol, beta-mercaptoethanol, dinitrothiobenzene), detergents, and acids. Biological agents, such as proteases, may also act as denaturants.

In some embodiments, the methods comprise

- a) providing a protein (I) comprising three cysteine residues as disclosed herein and

- b) contacting said protein with a trivalent thiol-reactive cross-linker as disclosed herein such that the linker forms covalent bonds with each of the three cysteine residues resulting in a cross-linked protein (II), wherein the cross-linked protein (II) has an increased Tm and/or an increased resistance to denaturation as compared to the protein lacking crosslinking (I).

The methods disclosed herein comprise providing a protein which comprises three cysteine residues. The protein is contacted with a trivalent thiol-reactive cross-linker such that the linker forms covalent bonds with each of the three cysteine residues. As described in the examples, bicyclization resulted in a stronger stabilization of protein tertiary structure compared to monocyclization, while still retaining protein function. In some embodiments, the three cysteine residues which are cross-linked are endogenous to the protein. However, for most proteins one or more of the three cysteine residues will be introduced into the protein. The introduction of cysteine residues may be accomplished by any method known to a skilled person, e.g, by chemically synthesizing the modified protein or by introducing the one or more cysteines using recombinant DNA technology. In some embodiments, modified proteins can be cloned in expression vectors and expressed in cell culture by techniques well-known in the art. It is also apparent that the disclosure encompasses proteins having more than one of the linkages described herein. For example, a protein may be provided with six cysteine residues, wherein three of the cysteine residues are able to be cross-linked while the other three cysteine residues are able to be cross-linked. In such embodiments, the cross-linker may be the same or different for each set of three cysteine residues.

Short peptides generally do not retain their native conformation. In fact, many peptides are extremely flexible. Thus peptide stapling techniques have been developed as a mean to constrain short peptides in a particular conformation and reduce backbone flexibility (reviewed in, e.g., Lau et al. 2015 Chem Soc Rev 44:91-102). It is also useful in drug discovery to constrain short linear peptides in order to adopt new structures with novel activities. For example, Bashiruddin et al. (2015 Bioorganic Chemistry 61: 45-50) describes a method of ribosomally synthesizing fused tricyclic peptides. The technology is used for generating libraries of peptides having new structures which can then be used, for example, in screening for bioactivity in particular to identify new functions. In this method, peptides are translated with an N-terminal ClAc group, one Cys residue at the second position and three more arbitrarily spaced downstream Cys residues, followed by the addition of TBMB. Bashiruddin et al. only tests short peptides having a length of less than 40 amino acids.

Chen et al. (2012 ChemBioChem 13: 1032-1038) is also concerned with peptide libraries for screening for high-affinity ligands. A 17-amino acid peptide was completely reduced with TCEP and incubated with different linker compounds dissolved in acetonitrile (an organic solvent). The authors conclude that combining different linkers with random peptide libraries could be a strategy for generating libraries of structurally highly diverse macrocyclic proteins.

Unlike peptide stapling which structurally reinforces a short string of amino acids in a particular secondary structure or loop conformation, the presently disclosed methods improve the stability of a protein tertiary structure. The disclosure concerns proteins having at least 70 amino acids. In some embodiments, the proteins have at least 80 amino acids or even at least 100 amino acids. Unlike the simple structure of a peptide, the multiple secondary structures of a protein fold to form a more complex three-dimensional structure. Preferably, the methods relate to proteins having at least two distinct secondary structures. Preferably, the protein provided in the method is a folded protein or rather a protein that has not been denatured before cross-linking. In one aspect the present invention provides methods for stabilizing the ‘natural’ folding or structure of a protein. In preferred embodiments, the methods increase the stability of the (natural) tertiary structure of a protein as compared to the non-cross-linked protein.

As used herein, “secondary structure” of a protein are defined by the patterns of hydrogen bonds between backbone amino and carboxyl groups. Alpha helices, beta sheets, beta turns and omega loops are exemplary secondary structures in proteins. As used herein, “tertiary structure” refers to the three dimensional shape of a protein.

Protein structure prediction techniques are well-known in the art and include homology modeling and threading, as well as more advanced methods that utilize neural networks, hidden Markov models and support vector machines. In addition, the tertiary structure of a protein can be determined by known-methods such as X-ray crystallography or nuclear magnetic resonance (NMR) studies. Publicly available software such as the Rosetta software can also be used for proteins structure prediction and to design new structures. See Voet, Pratt, Voet: Principles of Biochemistry, 2017 Chapter 6 Proteins: Three-Dimensional Structure for a review on protein folding and secondary structure; structure prediction and determining protein structure.

Preferably, the three cysteines for cross-linking are located in at least two distinct secondary structures. For example, the first cysteine may be located in a first alpha helix and the second cysteine may be located in a second alpha helix. The third cysteine may be located in either the first or second alpha helix or in a further secondary structure. Such methods have the advantage that the cross-linking increases the stability between at least two secondary structures. More preferably, the three cysteines for cross-linking are located in at least three distinct secondary structures.

The three cysteine residues are suitably located within the protein so that the cross-linker disclosed herein can form covalent bonds with each of the three cysteine residues. Preferably, the cysteine residues are separated in the primary amino acid sequence by at least 3 amino acids, while still being in spacial proximity. Preferably, the alpha-C atoms of the three cysteine residues form a triangle with side lengths between 6 to 23 Angstrom. Preferably, the cysteines are facing the same side of the protein.

Design principles known in the art of peptide stapling may be considered when introducing one or more cysteine residues into the protein for cross-linking. For example, it is known that amino acid residues of a peptide which lie on the same face of an alpha helix can be covalently joined or “stapled”. The spacing of such residues is generally, i,i+4, i+7, i+11, i+12, i+14 and i+15. In order to promote the stability of a beta-sheet, the spacing of residues to be stapled is generally, i, i,i+2, i+4, i+6, i+8, i+10, etc. The staple imparts rigidity, and reinforces the desired secondary structure of the peptide.

It is preferred that the preferred positions for the three cysteine residues are not buried or core positions. “Buried position” as used herein refers to positions that are in the interior of a protein and/or which are inaccessible or nearly inaccessible to solvent. The accessible surface area of a protein can be determined by a number of different prediction methods (see, e.g., Zheng, et al., Proteins: Structure, Function, and Bioinformatics. 2004;57:558–564; and Faraggi et al., Proteins. 2014 Nov; 82(11): 3170–3176). Preferably, the three cysteine residues are located on the surface of the tertiary structure and are not involved in binding (e.g., ligand binding, substrate recognition).

In some embodiments, the protein comprises at least a fourth cysteine residue, which is not cross-linked as a result of the method. Such methods are particularly useful when the protein comprises a cysteine residue that has a biological role, e.g., in a binding domain or enzymatic active site. As described in the examples, application of the present methods to a modified SrtA polypeptide surprisingly resulted in the cross-linking of three recombinantly introduced cysteine residues, while an endogenous cysteine residue which is crucial for enzymatic activity was not cross-linked.

One of the advantages of the methods disclosed herein is that the cross-linking does not rely on non-naturally occurring amino acids. In some embodiments, the protein does not comprise non-naturally occurring amino acids. The term “non-naturally occurring amino acid” includes amino acids that are different from naturally occurring amino acids in their side chain functionality. Naturally occurring amino acids include the 20 common amino acids: alanine, arginine, glycine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, serine, threonine, histidine, lysine, methionine, proline, valine, isoleucine, leucine, tyrosine, tryptophan, phenylalanine; as well as pyrolysine and selenocysteine.

In embodiments where non-endogenous cysteine residues are introduced into the protein, the effects of the modifications can be determined in, e.g., biological activity assays. For example, if the protein is an enzyme, the enzymatic activity of the protein having one or more cysteines introduced can be measured. While some loss of enzymatic activity is acceptable, modifications which significantly reduce enzymatic activity should be avoided. Similar assays can be performed to determine the effects of e.g., binding (e.g., affinity and specificity) and protein activity (e.g., downstream signaling). In vitro screening methods to measure the biological activity of proteins are well-known.

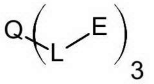
The present disclosure relates to trivalent thiol-reactive cross-linkers for increasing protein stability, as disclosed herein. As used herein, the term “cross-linker” refers to a reagent capable of chemically linking molecules, for example proteins, by one or more covalent bonds. The crosslinking reagents are “trivalent thiol-reactive”, i.e., they contain three reactive ends that are capable of attaching to a sulfhydryl group, e.g., a thiol side chain in cysteine. Preferably, the thiol-reactive end of the cross-linker comprises an electrophile. Preferably, the cross-linker is homo-trifunctional, or rather each thiol reactive end has the same functional group. Preferably, the cross-linker has a C3-symmetric core. This has the advantage that only one form of (tri)cross-linked protein will be formed. Trivalent thiol-reactive cross-linkers are known in the art (see, e.g., 26-29). However, such cross-linkers have not been described for being able to increase the stability of proteins having at least 70 amino acids, while maintaining protein function (in particular binding activity or enzymatic function).

In some embodiments, the trivalent thiol-reactive cross-linker comprises a fluorophore or an affinity handle. Suitable fluorophores are well-known in the art and include Alexa Fluor 350, Alexa Fluor 405, AMCA, Marina Blue dye, and Cascade Blue dye (available from Invitrogen). Affinity handles refers to molecules that can be used for detection and/or purification. Suitable affinity handles are known in the art and may include an antibody, a double-stranded DNA sequence, modified nucleic acids and nucleic acid mimics such as peptide nucleic acids, locked nucleic acids, phosphorodiamidate morpholino oligomers (PMO), a ligand, a receptor, a peptide, or a small molecule for which a cognate binding agent is readily available. Suitable affinity tags are peptide ‘tags’ such as polyhistidine, Calmodulin, S-tag, SBP-tag, Strep-tag, V5, FLAG, HA and Myc tags. Other suitable affinity tags are well-known in the art.

The cross-linking reaction is carried out under conditions known in the art. See, e.g., Mattson et al. Molecular Biology Reports 1993, Volume 17:pp 167–183; Paramelle et al. Proteomics 2013 13:438–456. In general, the reaction is carried out at a pH between 6-8 and at a temperature of between 4-40°C. Optionally, the efficiency and/or specificity of the cross-linking reaction can be determine, e.g., using MS. In a further aspect the present invention provides methods to stabilize the ‘natural’ folding or structure of a protein. Preferably, the reaction is carried out under conditions which do not disrupt the tertiary structure of the protein. Some cross-linkers require organic solvents for solubility. The presence of organic solvents may lead to the denaturation of the protein. In preferred embodiments, the reaction is carried out without the use of organic solvents. Preferred cross-linkers are those which do not require organic solvents for solubility.

In some embodiments, the methods further comprise determining the stability of the cross-linked protein. For example, the thermal and/or chemical stability can be determined as described herein and compared to the protein that has not been cross-linked. The biological activity of cross-linking can also be determined.

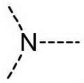
The present disclosure further provides cross-linkers having the formula (I):



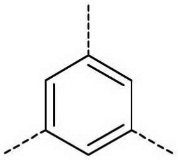
(I)

wherein

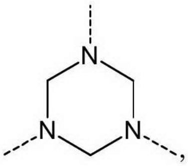
Q is a core structure selected from the group consisting of



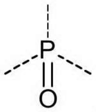
,



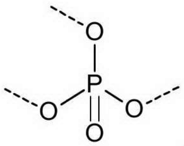
,



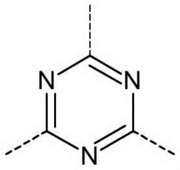
,



,



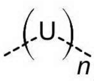
, and



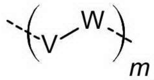
,

each dashed line in Q indicating a site where Q is bound to L,

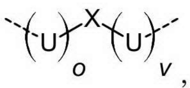
each L is a linker independently selected from the group consisting of



,



, and



,

wherein

each U is independently selected from CH_2 and CF_2,

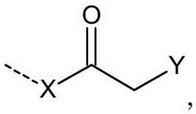
V is CH_2

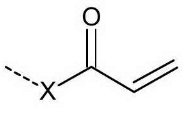
W is CF_2

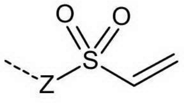
X is NR, NH or O,

wherein R is a fluorophore or affinity handle,
n is an integer in the range of 2-8,
m is an integer in the range of 1-4
o is 2 or 3, and
v is 2 or 3,
each dashed line in L indicating a site where L is bound to Q or E,

each E is an electrophile independently selected from the group consisting of



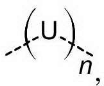
,


, and


,

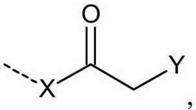
wherein
each X is independently selected from NH and O,
Y is selected from F, Cl, Br, Tos ($O-SO_2-C_6H_4-CH_3$), and Mes ($O-SO_2-CH_3$),
Z is CH_2, $NH-C(O)-CH_2$, or $O-C(O)-CH_2$
each dashed line in E indicating a site where L is bound to L.

Preferably, L is



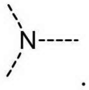
, U is CH_2, preferably n is 2 or 3; and/or

E is

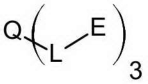


, preferably wherein X is NH and Y is F, Cl, or Br, preferably Cl.

While not wishing to be bound by theory, it is believed that trivalent cross-linkers having a non-hydrophic core (i.e, Q) are better suited to the cross-linking of proteins. In the known peptide stapling techniques which employ tris-electrophiles, cross-linkers with an aromatic core are generally used. In these cases, the core structures serve as the hydrophobic core to align non-polar amino acid side chains in its proximity. However, in the present preferred methods, a non-aromatic cross-linker will be located on the surface of the protein. Preferably, Q is



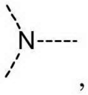
The present disclosure further provides cross-linkers having formula II:



(II)

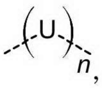
wherein

Q is

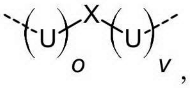


each dashed line in Q indicating a site where Q is bound to L,

each L is a linker independently selected from the group consisting of



and



wherein

each U is independently selected from CH_2 and CF_2,

V is CH_2

W is CF_2

X is NR, NH or O,

wherein R is a fluorophore or affinity handle, n is an integer in the range of 2-8,

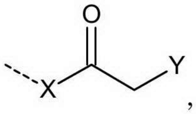
m is an integer in the range of 1-4

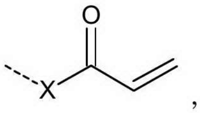
o is 2 or 3, and

v is 2 or 3,

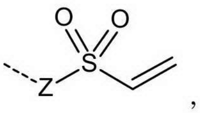
each dashed line in L indicating a site where L is bound to Q or E,

each E is an electrophile independently selected from the group consisting of





and



wherein

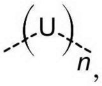
each X is independently selected from NH and O,

Y is selected from F, Cl, Br, Tos ($O-SO_2-C_6H_4-CH_3$), and Mes ($O-SO_2-CH_3$),

Z is CH_2, $NH-C(O)-CH_2$, or $O-C(O)-CH_2$

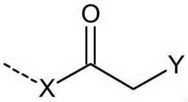
each dashed line in E indicating a site where E is bound to L.

Preferably, L is



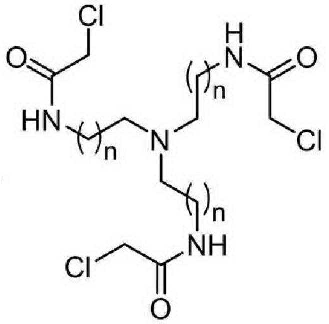
, U is CH_2, preferably n is 2 or 3; and/or

E is



, preferably wherein X is NH and Y is F, Cl, or Br, preferably Cl.

In a preferred embodiment, the cross-linker is Formula III:



wherein n is 1 or n is 2. As demonstrated in the examples, such cross-linkers do not require organic solvents for solubility.

The cross-linkers described above are particularly useful for reacting with thiol groups. As such they can be used to cross-link cysteine residues and can be used as the trivalent thiol-reactive cross-linker, e.g., in the methods described herein.

The disclosure further provides a stabilized protein obtainable by a method as disclosed herein. In one embodiment, the disclosure provides a protein (as disclosed herein, e.g., comprising at least 70 amino acids and comprising at least three cysteine residues), wherein each of the three cysteine residues is covalently bonded to a trivalent thiol-reactive cross-linker. Preferably, the protein is cross-linked by a cross-linker having a formula of (I), (II), or (III). In an exemplary embodiment, the protein is a Sortase A (SrtA) polypeptide. Preferably, the SrtA is cross-linked with formula (III).

SrtA is a transpeptidase belonging to the Sortase family of prokaryotic enzymes. It is an important biomolecular tool allowing specific labeling of proteins.[16-18] However, when higher temperatures or denaturants are required, labeling efficiency drops dramatically limiting the applicability of this enzyme. As described in the examples, the disclosure describes the generation of a SrtA polypeptide having a cysteine at positions 111, 149, and 177, with reference to amino acid position numbering of Staphylococcus aureus SrtA. Cross-linking this modified protein with a trivalent thiol-reactive cross-linker resulted in a 11.2 °C increase in melting temperature and an increased resistance to guanidinium hydrochloride, indicating an increase in stability. In a preferred embodiment, the SrtA polypeptide has the amino acid sequence:
GSHMQAKPQIPKDKSKVAGYIEIPDADIKEPVYPGPATPEQLNRGVSFAEENESLD
DQNISIAGHTFIDRPNYQFTNLKAAKKGSMVYFKVGNETRKYKMTSIRDVKPTDV
GVLDEQKGKDKQLTLITCDDYNEKTGVWEKRKIFVATEV (SEQ ID NO:1).

As described in the examples, cross-linked SrtA, as disclosed herein, can be used in protein labeling experiments under conditions where the wildtype (non-crosslinked) SrtA does not provide sufficient activity. The disclosure further provides the use of the cross-linked SrtA, as disclosed herein for protein/cell labeling, preferably in the presence of a chemical denaturing agent such as guanidinium hydrochloride.

In an exemplary embodiment, the protein is a KIX domain polypeptide. Preferably, the KIX domain is cross-linked with formula (III) or formula (IV). The disclosure provides KIX domain polypeptides comprising three cysteine residues is covalently bonded to a trivalent thiol-reactive cross-linker.

As used herein, "to comprise" and its conjugations is used in its non-limiting sense to mean that items following the word are included, but items not specifically mentioned are not excluded. In addition the verb “to consist” may be replaced by “to consist essentially of” meaning that a compound or adjunct compound as defined herein may comprise additional component(s) than the ones specifically identified, said additional component(s) not altering the unique characteristic of the invention.

The articles “a” and “an” are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.
The word “approximately” or “about” when used in association with a numerical value (approximately 10, about 10) preferably means that the value may be the given value of 10 more or less 1% of the value.

The invention is further explained in the following examples. These examples do not limit the scope of the invention, but merely serve to clarify the invention.

EXAMPLES

Example 1

Here we report a structure-based strategy for the stabilization of enzymes via post-translational modification of proteins entirely composed of proteinogenic amino acids. A library of biselectrophiles was used to staple a set of enzyme variants presenting pairs of accessible cysteine residues (Figure 1a). Based on the stabilization behavior of the resulting monocyclic proteins, a bicyclic enzyme was designed that shows greatly increased tolerance towards thermal and chemical denaturation.
We chose Staphylococcus aureus Sortase A (SrtA, aa 60–206) as the target for our stabilization efforts. To stabilize SrtA, we considered a crosslinking strategy that has previously been applied to constrain peptides[19-23] and involves the use of biselectrophiles that target pairs of cysteine residues. SrtA contains a single cysteine located in the active site which is crucial for its activity. Initially, we tested four electrophiles (1 – 4, Figure 1b) that have previously been used for the covalent modification of solvent exposed cysteines to seek for a functionality that does not react with the active site cysteine. Conditions suitable for preparative scale protein modification led to substantial modification of this crucial cysteine when incubated with the two most reative electrophiles maleimide (1) and 2-bromo¬acetamide (2) but not for 2-chloro¬acetamide (3) and acrylamide (4). As thiols tend to show reversible addition to acrylamides, we chose 2-chloroacetamide as the electrophile and designed

a set of biselectrophilic linkers with 8 – 17 bridging atoms (b1 – b6, Figure 2a) spawning a broad range of distances (up to 21 Å, Figure 5).

Next, suitable positions for the introduction of cysteine pairs in SrtA were selected aiming for a stabilization of the overall tertiary structure. We considered (i) surface residues not involved in substrate recognition and selected pairs (ii) that are located in two different secondary structure elements (iii) while still being in spacial proximity (distance <20 Å, based on NMR structure, PDB: 1ija). Based on these criteria, six SrtA variants (S1 – S6, Figure 2b, Figure 6) were designed, heterologously expressed in E. coli and purified. Subsequently, stapling reactions with all bis¬electrophiles were performed showing various degrees of conversion. Formation of the cyclization product was confirmed by MS and SDS-PAGE. While we observed high conversions for S1, S3, S4 and S6 with all crosslinks, S5 showed low efficiency with the shortest crosslink (b1). For S2, low yields were observed for all cross-linkers . After the reaction, protein samples were dialyzed to remove unreacted biselectrophiles.

Initially, the melting temperatures (Tm) of all unmodified and crosslinked variants (as obtained after dialysis) were determined via changes in tryptophan fluorescence (Figure 2c). Compared to SrtA (Tm = 59.4 °C), all non crosslinked variants show a lower thermal stability except for S3 (ΔTm = +2.9 °C). Enzyme cross-linking results in strong stabilization of the cyclic S3 versions (ΔTm ≥ +10.1 °C) while more moderate effects were observed for the remaining variants. The most stable versions per variant are S1 b1 (ΔTm = +2.8 °C), S2 b2 (ΔTm = +0.4 °C), S4 b3 (ΔTm = +4.4 °C), S5 b5 (ΔTm = +3.4 °C) and S6 b1 (ΔTm = +3.9 °C).

SrtA is a transpeptidase that recognizes a short peptide sequence (LPETG, Figure 2d), cleaves it and forms an acyl intermediate with its N-terminal fragment. The intermediate is then attacked preferably by the N-terminus of an oligo-glycine (Figure 2d) to form a new peptide bond. In absence of a suitable nucleophile, water will attack and hydrolyze the acyl intermediate (Figure 2d). To investigate transpeptidase activity, a previously reported probe system was applied in which a fluorophore/quencher pair is separated upon SrtA processing (Figure 7). For activity screening, we chose the hydrolysis reaction[24,25] at 65 °C, where wildtype SrtA shows strongly reduced performance (4 % residual activity, Figure 7). Relative to SrtA (vr = 1, Figure 2e), a number of crosslinked enzymes show increased activity.

Surprisingly, the thermostable cyclic versions of S3 provide reduced enzymatic activity (Figure 2e). In contrast, crosslinked versions of S4 and S5 show robust activity enhancements (>2-fold, light and dark red, Figure 2e). The overall highest increase in activity was observed for S4-b3, which is 3.4 fold more active than SrtA. Taken together, observed improvements in activity at 65 °C are moderate indicating that mono-cyclization may not be sufficient to convey enough stabilization of the tertiary structure.

To achieve a stronger stabilization of the protein tertiary structure, we aimed for a bicyclization of the enzyme. Notably, the two best performing SrtA variants S4 and S5 (light and dark blue, Figure 2b) share one variation site (aa 149). The simultaneous introduction of their cysteine substitutions (aa 111, 149 and 177) generates variant S7 (Figure 3a), which can form a bicyclic protein upon reaction with a triselectrophile. In analogy to the previously reported synthesis of bicyclic peptides[26-28] and mini-proteins[29], we selected a C3 symmetric core for our cross-linker, which we modified with three 2-chloroacetamide groups (t1, Figure 3b). Triselectrophile t1 was designed to provide 13 bridging atoms thereby lying between the preferred crosslink ranges for S4 (b3/b4: 10/11 atoms) and S5 (b5/b6: 14/17 atoms). The crosslinking reaction of S7 and t1 proceeds efficiently and provides the stapled enzyme S7-t1 (Figure 3b). Analytical HPLC/MS analysis indicates quantitative conversion of S7, clearly showing the formation of a product with the expected molecular weight. High-resolution MS analysis confirms the correct modification sites (cysteines 111, 149 and 177) also verifying the unmodified state of the active site cysteine after stapling (Figure 8). Investigating the thermal stability of S7-t1, we observed a greatly increased melting temperature (Tm = 70.6 °C, Figure 3c). This is 11.2 °C higher than the value of SrtA (Tm = 59.4 °C) and 6.8 °C higher than that of the most active mono-cyclic protein S4-b3 (Tm = 63.8 °C). Next, we determined the enzymatic activity of S7-t1 at 65 °C (Figure 3d) as described for the mono-cyclic versions. In line with its superior thermal stability, we observe a strongly increased enzymatic activity at 65 °C when compared to SrtA (8.7-fold) and to the most active mono-cyclic enzyme S4-b3 (2.6-fold, Figure 3d).

So far, we investigated enzyme activity under hydrolytic conditions using water as the nucleophile (hydrolysis, Figure 2d). Envisioning the application of S7-t1 for protein labeling, we next investigated its transpeptidation performance at 65 °C with the above described fluorescent probe but now in the presence of the nucleophile triglycine (transfer, Figure 2d). Using HPLC/MS as a readout (Figure 4a), we again observed only very weak substrate conversion with SrtA (dark grey) similar to a treatment without any enzyme (light grey). In the presence of S7-t1 (red, Figure 4a), the signal of the starting material (●), was greatly diminished and two new peaks appeared. Based on the MS, one peak was assigned to the C- (▲) and the other one to the N-terminal fragment (■), which appears to be ligated to triglycine. Importantly, a signal for the hydrolysis product (Dabcyl-QALPET) was not detected verifying the correct functionality of S7-t1. To assess if protein unfolding under elevated temperature is reversible, we compared the enzymatic activity of SrtA and S7-t1 at 37 °C before and after heating (85 °C Figure 11). Notably, the transpeptidase activity of both enzymes is not affected by the heating/cooling cycle indicating reversible unfolding.

In a next set of experiments, we determined the thermal activity profile of the transpeptidation reaction using again the fluorescent readout (Figure 4b). Between 37 °C and 55 °C, the enzymatic activity of SrtA (grey) and S7-t1 (red) is similar exhibiting only weak temperature dependence. Above 55 °C, both enzymes experience a loss in activity which is very severe for SrtA resulting in almost complete inactivation at 65 °C (Figure 4b). For S7-t1, the activity reduction is much smaller with a residual activity of 63 % (at 65 °C) and 27 % (at 70 °C) relative to 37 °C. Compared to SrtA, S7-t1 shows a ~10 °C increased tolerance towards thermal stress which correlates well with its +11.2 °C higher melting temperature. Enhanced thermal stability often goes in hand with a resistance towards denaturants such as guanidinium hydrochloride (GdnHCl). For that reason, the impact of GdnHCl on the transpeptidase activity was investigated (Figure 4c), revealing low dependence of SrtA and S7 t1 on the denaturant concentration up to 0.5 M. Between 0.75 and 1.5 M, S7-t1 is significantly more active than SrtA. Most notably at 1 M GdnHCl, SrtA does not show any detectable enzyme activity (vr < 1 %, Figure 4c) while S7-t1 still provides 40 % residual activity (compared to absence of GdnHCl). At higher GdnHCl concentrations (≥ 2 M) both enzymes lose their enzymatic activity.

So far, we have applied S7-t1 for the labeling of a short test peptide. Next, we were interested if S7-t1 is also useful for protein labeling in particular under conditions where wildtype SrtA does not provide sufficient activity. For that purpose, we chose α-

Synuclein (α-Syn) as the protein of interest. α-Syn comprises 140 amino acids and can form pathogenic fibrils which are associated with the onset of various neurodegenerative diseases including Parkinson's.[30] α Syn fibrils can be solubilized using GdnHCl.[31] We designed an α Syn version with a C-terminal SrtA-recognition motif to allow labeling. Following expression and purification, soluble α-Syn (A) was subjected to fibril formation and ultra-centrifugation.[31] Insoluble fibrils were washed and treated with buffer either lacking (B) or containing (C) GdnHCl (1 M).[31] When comparing the resulting soluble fractions (B and C) with the purified and soluble form of α-Syn (A, Figure 4d), we clearly observed re-solubilization only in the presence (C) but not in absence (B) of GdnHCl. To investigate protein labeling, these soluble samples (A, B, C) were incubated with either SrtA or S7-t1, and a fluorescent substrate (Figure 9). We then performed analysis via SDS-PAGE employing a fluorescence imager for the readout. For soluble α-Syn prior fibril formation (A) and therefore in absence of GdnHCl, SrtA and S7-t1 result in intense bands indicating efficient protein labeling (Figure 4e). As expected, under re-solubilization conditions lacking GdnHCl (B) and therefore also lacking soluble α-Syn, we did not observe any fluorescent signal (B, Figure 4e). On the contrary, for re-solubilization with GdnHCl (1 M), α-Syn labeling occurs but only with S7-t1 and not with wildtype SrtA (C). Notably, differences in the fluorescent band intensities for S7-t1 (A vs. C, Figure 4e) correlate well with the amount of α-Syn in the soluble fractions (A vs. C, Figure 4d) indicating good labeling efficiencies for S7 t1 in the presence of GdnHCl.

To assess the broader applicability of protein stabilization via bicyclization, we chose the KIX domain from the human CREB binding protein as a second target (Figure 5a). KIX is an adaptor domain with multiple protein binding partners that is composed of a central three α helix bundle (α1, α2, α3). The junction between this bundle and the C-terminal 310 helix (G1) is crucial for structural integrity (Figure 12a).[34] Thus, we focused on this area for tertiary structure stabilization searching for three positions suitable for cysteine incorporation. Based on our experiences during SrtA stabilization, the following guidelines were applied: (i) Solvent accessible residues were considered, that are (ii) located in three distinct secondary structures, while (iii) facing the same side of the protein and (iv) spanning a triangle with side lengths between 6 and 17 Å (Cα-Cα distance). Based on these criteria, we selected

H594, L599 and R646 for cysteine substitution resulting in KIX variant K1 (Figure 12a, Figure 12).

For crosslinking, we chose triselectrophile t1 (n = 2, Figure 3b) and a shorter version t2 (n = 1) since we noticed that the distances between the three variation sites in K1 (7.8, 10.0 and 11.5 Å, Figure 12) are shorter than in S7 (8.5, 12.4 and 15.7 Å, Figure 13d). The crosslinking with both triselectrophiles proceeds efficiently as confirmed by SDS-PAGE and HPLC-MS analysis (data not shown). To evaluate if crosslinking affects the tertiary structure, we compared the affinity of KIX and both bicyclic variants (K1-t1 and K1-t2) to its binding partner MLL. Using a fluorescence polarization assay, we observed similar binding affinities for KIX, K1-t1 and K1-t2 (Kd = 0.6, 0.9 and 0.9 μM, respectively, Figure 14). Then, we measured apparent melting temperatures for the three proteins (Figure 12b) to find a strongly increased thermal stability for K1-t1 and K1-t2 (ΔTm = +20.6 °C and +24.6 °C, respectively) when compared to KIX. Notably, both triselectrophiles have a similar stabilizing effect with the shorter crosslink t2 performing best. Based on these results, we were also interested to evaluate the effect of triselectrophile t2 on StrA variant S7. The crosslinking reaction proceeds efficiently resulting in bicyclic enzyme S7-t2 (Figure 15). Notably, we observe a similar thermal stabilization for S7-t2 (ΔTm = +11.5 °C, Figure 15) as for S7-t1 (ΔTm = +11.2 °C), indicating tolerance towards minor variabilities in the length of the crosslink.

In summary, we report a structure-based approach for the stabilization of enzymes that allows the installation of modular crosslinks into native proteins composed entirely of proteinogenic amino acids. We explored a series of mono-cyclized SrtA variants leading to the design of the bicyclic enzyme S7 t1 which exhibits greatly increased tolerance towards thermal and chemical denaturation. Importantly, S7 t1 proved efficient in labeling α-Syn in presence of 1 M GdnHCl. Under these conditions, wildtype SrtA did not show enzymatic activity. Even though, we did not encounter this problem with our SrtA variants, it is important to note that additional surface exposed cysteine residues can lead to undesired side products during the cyclization reaction. In such cases, it would be necessary to vary these cysteines (e.g. to serine) or if required for catalytic activity to block the active site during crosslinking. From our findings with SrtA, we derived guidelines for the bicyclization and stabilization of a protein and applied them to the KIX domain. A three cysteine KIX variant was designed and reacted with two different C3-symmentric triselectrophiles resulting in two bicyclic KIX versions both with greatly increased thermal stability. Overall, our approach facilitates a structure-based stabilization of recombinant proteins which are entirely composed of proteinogenic amino acids. The use of synthetic electrophiles for protein cyclization gives straight-forward access to diverse and tunable crosslink architectures. As an additional feature, we envision the use of crosslinking agents that allow the introduction of an additional functionality such as an affinity handle (e.g. for enzyme purification/recycling)[32] or a ligand for proximity-based sortase-mediated ligation.[33] Taken together, the presented protein stabilization technology holds the potential to give rapid access to novel stabilized enzymes providing the opportunity for a simultaneous incorporation of additional functions.

References

[1] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, Nature 2012, 485, 185–194.

[2] C. P. R. Hackenberger, D. Schwarzer, Angew. Chem. Int. Ed. 2008, 47, 10030–10074.

[3] A. J. Keefe, S. Jiang, Nat. Chem. 2012, 4, 59–63.

[4] J. Milton Harris, R. B. Chess, Nat. Rev. Drug Discov. 2003, 2, 214–221.

[5] M. T. Reetz, Angew. Chem. Int. Ed. 2013, 52, 2658–2666.

[6] C. Jost, A. Plückthun, Curr. Opin. Struct. Biol. 2014, 27, 102–112.

[7] T. J. Magliery, Curr. Opin. Struct. Biol. 2015, 33, 161–168.

[8] A. M. Chapman, B. R. McNaughton, Cell Chem. Biol. 2016, 23, 543–553.

[9] F. Agostini, J. S. Völler, B. Koksch, C. G. Acevedo-Rocha, V. Kubyshkin, N. Budisa, Angew. Chem. Int. Ed. 2017, 56, 9680–9703.

[10] N. Martínez-Sáez, S. Sun, D. Oldrini, P. Sormanni, O. Boutureira, F. Carboni, I. Compañón, M. J. Deery, M. Vendruscolo, F. Corzana, et al., Angew. Chem. Int. Ed. 2017, 56, 14963–14967.

[11] J. A. Camarero, T. W. Muir, J. Am. Chem. Soc. 1999, 121, 5597–5598.

[12] H. Iwai, A. Plu, FEBS Lett. 1999, 459, 166–172.

[13] C. Schoene, J. O. Fierer, S. P. Bennett, M. Howarth, Angew. Chem. Int. Ed. 2014, 53, 6101–6104.

[14] E. J. Moore, D. Zorine, W. A. Hansen, S. D. Khare, R. Fasan, Proc. Natl. Acad. Sci. 2017, 201708907.

[15] J. W. Chin, EMBO J. 2011, 30, 2312–2324.

[16] C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. M. Blom, H. L. Ploegh, Nat. Protoc. 2013, 8, 1787–1799.

[17] J. J. Bellucci, J. Bhattacharyya, A. Chilkoti, Angew. Chem. Int. Ed. 2015, 54, 441–445.

[18] L. Schmohl, D. Schwarzer, Curr. Opin. Chem. Biol. 2014, 22, 122–128.

[19] H. Jo, N. Meinhardt, Y. Wu, S. Kulkarni, X. Hu, K. E. Low, P. L. Davies, W. F. Degrado, D. C. Greenbaum, J. Am. Chem. Soc. 2012, 134, 17704–17713.

[20] A. Muppidi, K. Doi, S. Edwardraja, E. J. Drake, A. M. Gulick, H. G. Wang, Q. Lin, J. Am. Chem. Soc. 2012, 134, 14734–14737.

[21] P. Diderich, D. Bertoldo, P. Dessen, M. M. Khan, I. Pizzitola, W. Held, J. Huelsken, C. Heinis, ACS Chem. Biol. 2016, 11, 1422–1427.

[22] A. M. Spokoyny, Y. Zou, J. J. Ling, H. Yu, Y. S. Lin, B. L. Pentelute, J. Am. Chem. Soc. 2013, 135, 5946–5949.

[23] M. R. Jafari, L. Deng, P. I. Kitov, S. Ng, W. L. Matochko, K. F. Tjhung, A. Zeberoff, A. Elias, J. S. Klassen, R. Derda, ACS Chem. Biol. 2014, 9, 443–450.

[24] B. C. Chenna, J. R. King, B. A. Shinkre, A. L. Glover, A. L. Lucius, S. E. Velu, Eur. J. Med. Chem. 2010, 45, 3752–3761.

[25] K. B. Oh, S. H. Kim, J. Lee, W. J. Cho, T. Lee, S. Kim, J. Med. Chem. 2004, 47, 2418–2421.

[26] P. Timmerman, J. Beld, W. C. Puijk, R. H. Meloen, ChemBioChem 2005, 6, 821–824.

[27] C. Heinis, T. Rutherford, S. Freund, G. Winter, Nat. Chem. Biol. 2009, 5, 502–507.

[28] S. Chen, D. Bertoldo, A. Angelini, F. Pojer, C. Heinis, Angew. Chem. Int. Ed. 2014, 53, 1602–1606.

[29] B. Dang, H. Wu, V. K. Mulligan, M. Mravic, Y. Wu, T. Lemmin, A. Ford, D.-A. Silva, D. Baker, W. F. DeGrado, Proc. Natl. Acad. Sci. 2017, 201710695.

[30] M. R. Cookson, Mol. Neurodegener. 2009, 4, 9.

[31] R. Porcari, C. Proukakis, C. A. Waudby, B. Bolognesi, P. P. Mangione, J. F. S. Paton, S. Mullin, L. D. Cabrita, A. Penco, A. Relini, et al., J. Biol. Chem. 2015, 290, 2395–2404.

[32] C. B. Rosen, R. L. Kwant, J. I. MacDonald, M. Rao, M. B. Francis, Angew. Chem. Int. Ed. 2016, 55, 8585–8589.

[33] H. H. Wang, B. Altun, K. Nwe, A. Tsourkas, Angew. Chem. Int. Ed. 2017, 56, 5349–5352.

[34] R. N. De Guzman, N. K. Goto, H. J. Dyson, P. E. Wright, J. Mol. Biol. 2006, 355, 1005–1013.

Methods

1.1 Peptide synthesis and characterization

Peptide synthesis was performed manually on Fmoc-Rink Amid MBHA resin (Iris Biotech GmbH) according to standard Fmoc-based solid-phase peptide synthesis (SPPS) protocols. All Fmoc-protected amino acids were coupled using 4 eq. calculated to the initial amine-loading of the resin. The coupling conditions were 4 eq. of (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)¬dimethylamino-morpholino-carbenium hexafluorophosphate (COMU), 4 eq. of oxyma and 8 eq. of N,N-diisopropylethylamine (DIPEA) in dimethylformamide (DMF) for the first coupling reaction (20 min). For the second coupling (45 min) 4 eq. of benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) and 8 eq. of DIPEA in DMF were used. Fmoc-deprotection was accomplished treating the resin with a solution of 20 % piperidine in DMF for 15 min. Fluorescein isothiocyanate (FITC) was coupled using 3 eq. of its isomer and 6 eq. of DIPEA in DMF for 2 h twice. Peptides were cleaved from the resin after treatment with a TFA:H2O:TIPS (95:2.5:2.5) solution (2x 2 h) and precipitated with Et2O at -20 °C. After lyophilization in a Freezone 4.5-105 °C freeze drying system (Labconco), the peptides were dissolved in H2O:acetonitrile (1:1) and purified in a reversed-phase semi-preparative HPLC using a Nucleodur C18 reverse-phase column (10 x 125 mm, 110 Å, particle size 5 µm, Macherey-Nagel; solvent A: H2O + 0.1 % TFA; solvent B: acetonitrile + 0.1 % TFA; flow rate of 6 mL min-1). Obtained products were lyophilized.

Peptide identity and purity were confirmed by HPLC/ESI-MS analysis performed in a HPLC-MS system (Agilent Technologies) provided with a Zorbax Eclipse, XDB-C18 reverse-phase column (4.6 x 150 mm, particle size 5 µm, Agilent; solvent A: H2O + 0.1 % TFA; solvent B: acetonitrile + 0.1 % TFA; flow rate of 1 mL min-1). The FITC labeled peptides were quantified photometrically using a V-550 UV/VIS spectrophotometer (Jasco). For the GGGK-FITC peptide, absorbance at 494 nm was measured in a 100 mM sodium phosphate buffer (pH 8.5) and concentration calculated using the extinction coefficient ε(FITC)494 = 77000 M-1 cm-1. For the Dabcyl-QALPETGEK-FITC peptide, Dabcyl absorption at 494 nm was additionally taken into account (ε(Dabcyl)494 = 14000 M-1 cm-1).

1.2 Protein expression and purification

A modified pET28a(+) vector coding for Staphylococcus aureus SrtA aa 60-206 was provided by AG Musacchio (Max Planck Institute, Dortmund, Germany). Variants S1 – S7 were obtained either by sequence modification using site-directed mutagenesis (QuikchangeTM, Stratagene), restriction and ligation or in vivo cloning each resulting in the according N-terminal His6-tagged protein (Figure 6). These constructs were transformed into E. coli BL21 Gold (DE3) (Agilent Technologies). Transformants were used to inoculate a Luria Broth (LB) (50 µg mL-1 kanamycin) overnight pre-culture (incubated at 37 °C). This culture was used to inoculate a Terrific Broth (TB) culture (2 L) which was incubated at 37 °C until an OD600 of 0.7 was reached. Protein expression was induced by addition of 0.5 mM IPTG and the culture was incubated overnight at 25 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl, 2 mM TCEP, 10 % glycerol (v/v)) and disrupted in a microfluidizer. The cell lysate was cleared from cell debris by centrifugation (70000 rcf, 4 °C, 45 min). All subsequent purification steps were performed at 4 °C. SrtA and variants S1 – S7 were isolated from the supernatant via FPLC affinity chromatography (HisTrapTM Fast Flow Crude 5 mL, GE Healthcare). The column was washed with 5 CV washing buffer (50 mM Tris (pH 7.5), 150 mM NaCl, 2 mM TCEP, 5 % glycerol (v/v), 20 mM imidazole). Thrombin cleavage (5 U mg-1) was performed on column over night at 4 °C in Thrombin buffer (50 mM Tris (pH 8), 100 mM NaCl, 2.5 mM CaCl2, 1 mM DTT) resulting in target protein elution. To separate SrtA and variants from Thrombin enzyme size exclusion chromatography was performed (Aekta Pure, Column HiLoad 16/600 Superdex 75 pg, GE Healthcare in 20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP). The purified proteins were concentrated via ultra-filtration (Amicon, Merck, 10 kDa cut off) up to 25 mg mL-1, snap frozen and stored at -80 °C.

The coding sequence for an α-Synuclein (α-Syn) construct containing a C-terminal flexible linker, a SrtA recognition site as well as a His6-tag for affinity purification (Figure 9), was purchased as gene synthesis (Integrated DNA Technologies) and sub-cloned into a pET28a(+) vector via restriction (NcoI and XhoI) and ligation. For protein expression the vector was transformed into E. coli BL21 Gold (DE3). Transformants were used to inoculate an overnight LB pre-culture (50 µg mL-1 kanamycin, 37 °C). This culture was used to inoculate a kanamycin (50 µg mL-1) containing LB culture (2 L) which was incubated at 37 °C until an OD600 of 0.7 was reached. Protein expression was induced by addition of 0.5 mM IPTG and performed for 4 h at 37 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl) and disrupted by ultrasonication (4x 30 s cycles with 1 s on 2 s off, 40 % power, 0 °C). The cell lysate was cleared from cell debris by centrifugation (70000 rcf, 1 h). All follow up purification steps were performed at 4 °C. α-Syn was isolated from the supernatant via affinity chromatography (Aekta Pure, HisTrapTM Fast Flow Crude 5 mL, GE Healthcare). Washing was performed with lysis buffer containing 5 mM imidazole. α Syn was eluted off the column in an imidazole gradient (5 mM-500 mM). Protein containing fractions were pooled and dialyzed (Slide A Lyzer Dialysis Cassette, Thermo, 3.5 kDa cut off) overnight at 4 °C against PBS buffer (pH 7.4). Resulting pure α-Syn was concentrated via ultra filtration (Amicon, Merck, 3 kDa cut off) up to 5 mg mL-1 and subjected to fibril formation.

The coding sequences for KIX and K1 constructs (Figure 13) containing Gateway attB1 and attB2 (Thermo) attachment and a PreScission protease recognition site, were synthesized and purchased from Integrated DNA Technologies. Afterwards, coding regions were introduced into a pDONR201 vector (BP Clonase enzyme mix, Thermo). Subsequently, LR Clonase enzyme mix (Thermo) was utilized to introduce the coding region into a pGEX-4t-3 Gateway compatible destination vector. The resulting expression vector was transformed into E. coli BL21 Gold (DE3). Transformants were used to inoculate an overnight LB pre-culture (100 µg mL-1 ampicillin, 37 °C) and subsequently this culture was used to inoculated an ampicillin (100 µg mL-1) containing TB culture (2 L) which was incubated at 37 °C until an

OD600 of 1 was reached. Protein expression was induced by addition of 0.5 mM IPTG and performed overnight at 20 °C. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris (pH 7.4), 500 mM NaCl, 2 mM PMSF and 2 mM DTT) and disrupted using the microfluidizer. The cell lysate was cleared from cell debris by centrifugation (70000 rcf, 4 °C, 60 min). All subsequent purification steps were performed at 4 °C. KIX and K1 were isolated from the supernatant via affinity chromatography (Aekta Pure, GSTPrepTM FF 16/10, GE Healthcare). Washing was performed with wash buffer (50 mM Tris (pH 7.4), 100 mM NaCl, 2 mM DTT) until baseline (OD280) was reached. PreScission cleavage was performed on column over night at 4 °C in wash buffer. Resulting target protein was concentrated via ultra-filtration up to about 6 mg mL-1 (Amicon, Merck, 3 kDa cut off, r.t.). Subsequent size exclusion chromatography was performed (Aekta Pure, Column HiLoad 16/600 Superdex 75 pg, GE Healthcare in 25 mM HEPES (pH 7.4), 100 mM NaCl, 2 mM TCEP). The purified proteins were concentrated (Amicon, Merck, 3 kDa cut off, r.t.) up to 6 mg mL-1, snap frozen and stored at -80 °C.

All generated vector constructs were sequence proven by Sanger sequencing. All proteins where checked for their quality via SDS-PAGE.

1.3 α-Syn fibril formation and re-solubilization

For fibril formation, 4 mL of purified, soluble α-Syn (5 mg mL-1) were stirred at 37 °C, 1250 rpm for 4 days. The resulting suspension was ultracentrifuged (135000 rcf, 4 °C, 45 min.) in 600 μL aliquots. The protein content of the supernatant was quantified via Nanodrop (OD280) to monitor the efficiency of fibril formation. After removal of the supernatant, the pellet was washed extensively with SrtA buffer (4 times 500 μL, 20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP). For re-solubilization the fibril sample was treated either with (1 M) or without GdnHCl in SrtA buffer for 3 h at RT and gentle shaking. Supernatant was used for subsequent labeling.

1.4 α-Syn labeling reaction

Re-solubilized α-Syn fractions were diluted 1:6 with SrtA buffer (+/- 1 M GdnHCl) keeping the initial GndHCl concentration and supplemented with 2 mM GGGK-FITC and either 100 μM SrtA or S7-t1. The labeling samples were incubated for 16 h at 350 rpm and 37 °C. Samples were analyzed on a SDS-PAGE gel for fluorescence in a Gel Doc XR system (BioRad). Then the gel was coomassie stained (Figure 10).

1.5 Synthesis and characterization of bis- and triselectrophilic cross-linkers

To a solution of K2CO3 (33 mmol, 3.3 eq.) in H2O/DCM (2:3, 18 mL) at 0 °C, the corresponding diamine (10 mmol, 1 eq.) was added. The resulting mixture was allowed to cool down before the chloroacetyl chloride (22 mmol, 2.2 eq.) was added dropwise over a 1 h period at 0 °C. After completed addition, the ice bath was removed and the mixture was allowed to stir at room temperature overnight. The desired product was extracted three times with DCM. Subsequently, the organic layer was washed with brine, dried over Na2SO4, filtrated and concentrated under reduced pressure. Product identity, was confirmed by NMR. For the triselectrophilic cross-linker t1 the same protocol was used with adjusted equivalents. In any crosslinking reaction, freshly prepared cross-linkers must be used.

N,N'-bis(chloroacetyl)-1,2-ethylenediamine (b1). 1H NMR (400 MHz, DMSO-d6) δ 8.27 (bs, CONH, 2H), 4.04 (s, CH2, 4H), 3.17 (m, CH2, 4H).

N,N'-bis(chloroacetyl)-1,3-propylenediamine (b2). 1H NMR (400 MHz, DMSO-d6) δ 8.20 (m, CONH, 2H), 4.04 (s, CH2, 4H), 3.10 (td, J = 6.9, 5.7 Hz, CH2, 4H), 1.57 (p, J = 6.9 Hz, CH2, 2H).

N,N'-bis(chloroacetyl)-1,4-butanediamine (b3). 1H NMR (400 MHz, DMSO-d6) δ 8.19 (m, CONH, 2H), 4.02 (s, CH2, 4H), 3.08 (m, CH2, 4H), 1.41 (m, CH2, 4H).

N,N'-(oxybis(ethane-2,1-diyl))bis(2-chloroacetamide) (b4). 1H NMR (400 MHz, CDCl3) δ 6.92 (bs, CONH, 2H), 4.07 (s, CH2, 4H), 3.59 (m, CH2, 4H), 3.53 (m, CH2, 4H).

N,N'-((ethane-1,2-diylbis(oxy))bis(ethane-2,1-diyl))bis(2-chloroacetamide) (b5). 1H NMR (400 MHz, DMSO-d6) δ 8.23 (m, CONH, 2H), 4.06 (s, CH2, 4H), 3.52 (m, CH2, 4H), 3.44 (t, J = 5.8 Hz, CH2, 4H), 3.25 (q, J = 5.8 Hz, CH2, 4H).

N,N'-(((oxybis(ethane-2,1-diyl))bis(oxy))bis(ethane-2,1-diyl))bis(2-chloroacetamide) (b6). 1H NMR (400 MHz, DMSO-d6) δ 8.23 (m, CONH, 2H), 4.06 (s, CH2, 4H), 3.52 (m, CH2, 8H), 3.44 (t, J = 5.8 Hz, CH2, 4H), 3.25 (q, J = 5.8 Hz, CH2, 4H).

N,N',N''-(nitrilotris(propane-3,1-diyl))tris(2-chloroacetamide) (t1). 1H NMR (400 MHz, DMSO-d6) δ 8.23 (t, J = 5.6 Hz, CONH, 3H), 4.03 (s, CH2, 6H), 3.10 (q, J = 6.6 Hz, CH2, 6H), 2.34 (t, J = 6.9 Hz, CH2, 6H), 1.52 (p, J = 7.1 Hz, CH2, 6H).

1.6 Protein modification with electrophiles

The reactivity of four different electrophiles with wildtype SrtA was evaluated. Thus, 50 µM SrtA was incubated with 2 mM electrophile (acrylamide 1, 2-bromoacetamide 2, 2-chloroacetamide 3 or 4 maleimidobutyric acid 4) in crosslinking buffer (50 mM

HEPES (pH 8.5), 150 mM NaCl, 5 mM CaCl2 and 2 mM TCEP) at 35 °C and 350 rpm for 24 h. The reactions were analyzed by MS.

Protein variants S1 – S6 were diluted to 50 µM in reaction buffer (50 mM HEPES (pH 8.5), 150 mM NaCl, 5 mM CaCl2 and 2 mM TCEP) and incubated with 0.5 mM biselectrophilic cross-linkers (b1 – b6, 50 mM in DMSO) at 35 °C and 350 rpm for 24 h. To stop the reaction, solutions were concentrated by ultra filtration (Amicon Ultra centrifugal filters, 0.5 mL, Merck, 10 kDa cut off) and washed 5 times with SrtA buffer (20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP) in order to remove the low molecular weight biselectrophilic cross-linkers, to exchange buffer and to concentrate final crosslinked protein variants. The resulting proteins were snap frozen and stored at -80 °C. For the crosslinking of variant S7 with the triselectrophilic cross-linker t1, the protocol described above was applied only changing the concentration of t1 to 1 mM instead of 0.5 mM (Figure 3b).

1.7 Measurement of melting temperature (Tm)

Each protein was diluted to 75 µM in 20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP. Temperature was scanned (20 – 95 °C) in a Prometheus NT.48 (NanoTemper Technologies) at a heating rate of 1 °C min-1 using an excitation power of 35 %. The ratio of fluorescence intensities at 350 nm and 330 nm (F350/F330) was plotted against the temperature and Tm-values were determined using the Nanotemper technologies protocols.

1.8 Enzymatic hydrolysis assay

Hydrolysis of probe Dabcyl-QALPETGEK-FITC (Figure 7) by StrA and variants provide the cleavage product GEK-FITC which exhibits increased fluorescence. Changes in fluorescence were monitored using a Real Time PCR system (StepOnePlusTM Real-Time PCR System, Applied Biosystems) by measuring fluorescence (FAM channel, 520 nm) at the given temperature (in 10 min steps over 16 h). Enzymes were diluted to 20 µM in SrtA buffer (20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP) supplemented with Tween (final concentration 0.01 %). Resulting enzyme solutions were mixed 1:1 with a 20 µM solution of the peptidic probe (Dabcyl-QALPETGEK-FITC) in the same buffer (final volume 20 µL, 10 µM probe, 10 µM enzyme). A sample without enzyme was used as blank. Fluorescence readout was background substracted and plotted against time. The slope of the linear part of the curve was determined (v) as a measure of enzymatic activity. Subsequently, v-values were divided by v(SrtA) to obtain relative activities.

1.9 Enzymatic transpeptidation assay

Fluorescent readout

This assay was performed analogously to the hydrolysis assay described above, but in presence of 2.5 mM triglycine (G3, Sigma-Aldrich). To perform the thermal activity profile of SrtA and S7-t1, transpeptidation activity was determined at various temperatures (37 °C, 45 °C, 55 °C, 60 °C, 65 °C, 70 °C, 75 °C). Triplicates were measured, averaged and plotted with error bars (1 σ). Statistical significance was evaluated by an unpaired t-test (GraphPad). We considered p-values < 0.05 as statistically significant (ns: not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, Figure 4b).

HPLC-MS monitoring of transpeptidation activity

Reactions were performed at 65 °C for SrtA and S7-t1. Reaction conditions: 50 μM POI (SrtA or S7-t1); 10 μM peptidic probe (Dabcyl-QALPETGEK-FITC); 2.5 mM G3 in SrtA buffer (20 mM HEPES (pH 7.5), 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP); 65 °C, 12 h, 350 rpm. The reaction mixture was quenched by addition of 1 % TFA. The products of these reactions were analyzed by HPLC-MS (Figure 4a).

Transpeptidation activity in the presence of GdnHCl

Transpeptidation activity efficiency in the presence of the denaturing agent GdnHCl was evaluated for SrtA and S7-t1. Transpeptidation activity at fixed POI (10 μM), peptidic probe (10 μM) and G3 (2.5 mM) concentrations and at fix temperature (37 °C) was measured at several GdnHCl concentrations (0 M, 0.5 M, 0.75 M, 1.0 M, 1.5 M and 2.0 M). Triplicates were measured, averaged and plotted with error bars (1 σ). Statistical significance was evaluated by an unpaired t-test (GraphPad). We considered p-values < 0.05 as statistically significant (ns: not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, Figure 4c).

1.10 Folding reversibility of SrtA and S7-t1

SrtA and S7-t1 were diluted to 20 μM in 20 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM CaCl2, 2 mM TCEP and 0.01% Tween 20. These solutions were heated from r.t. to 85 °C over 30 min in a ThermoMixer (HTA-BioTec), and then cooled down to r.t. over 60 min. Afterwards, the proteins were evaluated for transpeptidation activity at 37 °C together with freshly prepared solutions of non-preheated proteins.

1.11 HPLC-coupled High-resolution Mass spectrometry

Unmodified S7 and bicyclic S7-t1 were incubated first with 1 mM DTT then with 5.5 mM iodoacetamide, denatured in 8 M Urea and finally digested. First, LysC (WakoTM, Osaka, Japan) was used for 3 hours (protein-to-enzyme ratio 50:1) and after dilution with 4 volumes of 50 mM ammonium bicarbonate (AMBIC, pH 8.3) to a final concentration of 2.0 M urea, peptides were digested overnight at 37 °C with sequencing-grade modified trypsin (PromegaTM, Madison, WI; protein-to-enzyme ratio 50:1). Peptides were desalted on C18 stage tips and 100-300 ng of peptide were separated with a PepMap100 RSLC C18 nano-HPLC column (2 µm, 100 Å, 75 IDx25 cm, nanoViper, Dionex, Germany) on an UltiMate™ 3000 RSLCnano system (ThermoFisher Scientific, Germany) using a 65 min gradient from 5 – 60% acetonitrile with 0.1% formic acid and then directly sprayed via a nano-electrospray source (Nanospray Flex Ion Scource, Thermo Scientific) in a Q ExactiveTM Hybrid Quadrupole-Orbitrap Mass Spectrometer HF (ThermoFisher Scientific). For coupling of the nano-HPLC to the Quadrupole-Orbitrap Mass Spectrometer, a standard coated SilicaTip (ID 20 µm, Tip-ID 10 µm, New Objective, Woburn, MA, USA) was used. The Q ExactiveTM HF was operated in a data-dependent mode acquiring one survey scan and subsequently ten MS/MS scans. MS spectra were acquired with a mass range from 300 to 1650 m/z with a resolution of 70000, followed by up to then high energy collision dissociation (HCD) MS/MS scans at a resolution of 17500. Resulting raw files were processed with the MaxQuant software (version 1.5.2.18) including the Andromeda search algorithm searching against S7 using deamidation (de, for Asn and Gln), oxidation (ox, for Met), carbamidomethylation (ca, for Cys) and t1 remnant (cl, $C_2H_2O[-H]$ for Cys) as variable modifications. The mass accuracy for full mass spectra was set to 20 ppm for the first and to 4.5 ppm for the second search and to 20 ppm for MS/MS spectra. Two miscleavages were allowed. A false discovery rate cut off of 1% was applied at the peptide and site decoy fraction.

1.12 Fluorescence Polarization Assay (FP)

Binding of K1-t1 and K1-t2 to mixed lineage leukemia (MLL) transcription factor was evaluated using the FITC-labeled peptide FITC-O2Oc-GNILPSDI(Nle)DFVLKNTP-NH2. This sequence is derived from MLL and will be refereed as MLL peptide throughout. A 40 nM solution of MLL peptide in 25 mM HEPES (pH 7.4), 100 mM NaCl and 2 mM TCEP was prepared. A 3-fold dilution of KIX, K1-t1 and K1-t2 was performed in 14 steps on a 384-well-plate (black, round bottom, Corning) using the same buffer. 5 µL of 40 nM MLL peptide solution was then added, and the final 20 µL solution was incubated at r.t.. A final protein range of 70 µM – 0 µM was used. After 1h-incubation, fluorescence polarization was measured using a Spark 20M plate reader (Tecan) with λ(ex) = 485 nm and λ(em) = 525 nm. Kd were determined by applying nonlinear regression analysis of dose-response curves in GraphPad Prism software.

Example 2

Aldehyde dehydrogenases (ALDHs) have been applied for their highly chemoselective oxidation of aldehyde moieties to carboxylic acids on many different substrate molecules [T. Knaus, V. Tseliou, L. D. Humphreys, N. S. Scrutton, F. G. Mutti, Green Chemistry 20181]. The melting temperature is slightly above 47 °C, meaning that only after two to four hours, the catalytic activity of ALDH is significantly reduced. Cross-linked ALDH can help to improve thermal stability of the enzyme and increase its “longevity” during biocatalytic oxidation. ALDH from bovine lens (ALDH-Bov) forms a homodimer and/or -tetramer, of which the quarternary structure is available. To crosslink three monomers of the multimer, an ALDH-Bov polypeptide is designed having cysteines at positions 73, 414, 499, to crosslink two monomers of the multimer, an ALDH-Bov polypeptide is designed having cysteines at positions 72, 238, 448 with reference to amino acid position numbering of ALDH-Bov polypeptide having the amino acid sequence:

MSSSAMPDVPAPLTNLQFKYTKIFINNEWHSSVSGKKFPVFNPATEEKLCEVEEG
DKEDVDKAVKAARQAFQIGSPWRTMDASERGRLLNKLADLIERDHLLLATMEAM
NGGKLFSNAYLMDLGGCIKTLRYCAGWADKIQGRTIPMDGNFFTYTRSEPVGVCG
QIIPWNFPLLMFLWKIGPALSCGNTVVVKPAEQTPLTALHMGSLIKEAGFPPGVVN
IVPGYGPTAGAAISSHMDVDKVAFTGSTEVGKLIKEAAGKSNLKRVSLELGGKSPC
IVFADADLDNAVEFAHQGVFYHQGQCCIAASRLFVEESIYDEFVRRSVERAKKYVL
GNPLTPGVSQGPQIDKEQYEKILDLIESGKKEGAKLECGGGPWGNKGYFIQPTVFS
DVTDDMRIAKEEIFGPVQQIMKFKSLDDVIKRANNTFYGLSAGIFTNDIDKAITVSS
ALQSGTVWVNCYSVVSAQCPFGGFKMSGNGRELGEYGFHEYTEVKTVTIKISQKN
S.

Such crosslinked ALDH-Bov polypeptides are expected to have increased thermal stability.

Claims

1. A method for increasing the stability of a protein, wherein the protein comprises at least 70 amino acids, said method comprising:

- a) providing a protein comprising three cysteine residues and
- b) contacting said protein with a trivalent thiol-reactive cross-linker such that the linker forms covalent bonds with each of the three cysteine residues, wherein the method increases the stability of the tertiary structure of the protein as compared to the non-cross-linked protein.

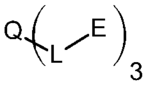
2. The method of claim 1, wherein step a) comprises modifying a protein to introduce one or more of the three cysteine residues.

3. The method of claim 1 or 2, wherein said protein comprises at least a fourth cysteine residue and wherein said method does not result in the formation of a covalent bond between the fourth cysteine and the cross-linker.

4. The method of claim 3, wherein the protein is an enzyme and the fourth cysteine is part of the enzymatic active site.

5. The method of any one of the preceding claims, wherein the cross-linker has C_3 symmetry.

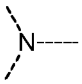
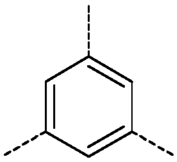
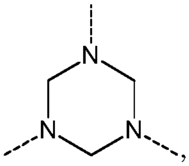
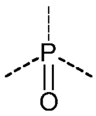
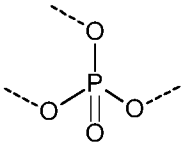
6. The method of any one of the preceding claims, wherein the cross-linker has formula (I):

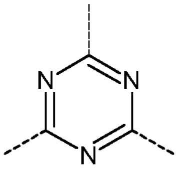


(I)

wherein

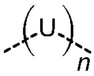
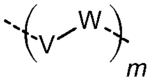
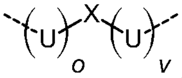
Q is a core structure selected from the group consisting of


, 
, 
, 
, 
, and


,

each dashed line in Q indicating a site where Q is bound to L,

each L is a linker independently selected from the group consisting of


, 
, and 
,

wherein

each U is independently selected from CH_2 and CF_2,

V is CH_2

W is CF_2

X is NR, NH or O,

wherein R is a fluorophore or affinity handle,

n is an integer in the range of 2-8,

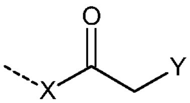
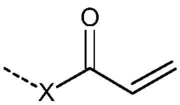
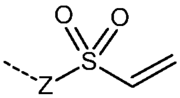
m is an integer in the range of 1-4

o is 2 or 3, and

v is 2 or 3,

each dashed line in L indicating a site where L is bound to Q or E,

each E is an electrophile independently selected from the group consisting of


, 
, and 
,

wherein

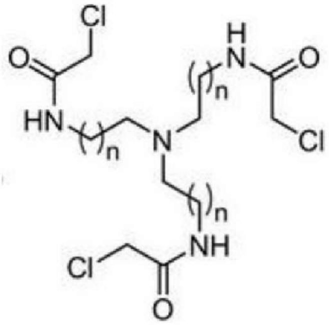
each X is independently selected from NH and O,

Y is selected from F, Cl, Br, Tos ($O{-}SO_2{-}C_6H_4{-}CH_3$), and Mes ($O{-}SO_2{-}CH_3$),

Z is CH_2, $NH{-}C(O){-}CH_2$, or $O{-}C(O){-}CH_2$

each dashed line in E indicating a site where E is bound to L.

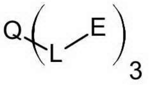
7. The method of claim 6, wherein the cross-linker has formula (III):



wherein n is 1 or n is 2.

8. A stabilized protein obtainable by a method according to any one of the preceding claims.

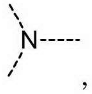
9. A trivalent thiol-reactive cross-linker having formula II:



(II)

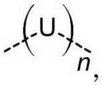
wherein

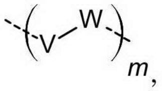
Q is



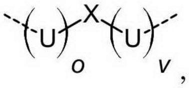
each dashed line in Q indicating a site where Q is bound to L,

each L is a linker independently selected from the group consisting of





and



wherein

each U is independently selected from CH_2 and CF_2,

V is CH_2

W is CF_2

X is NR, NH or O,

wherein R is a fluorophore or affinity handle, n is an integer in the range of 2-8,

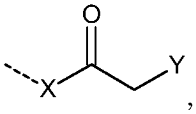
m is an integer in the range of 1-4

o is 2 or 3, and

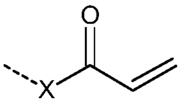
v is 2 or 3,

each dashed line in L indicating a site where L is bound to Q or E,

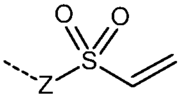
each E is an electrophile independently selected from the group consisting of



,



, and



,

wherein

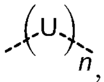
each X is independently selected from NH and O,

Y is selected from F, Cl, Br, Tos ($O-SO_2-C_6H_4-CH_3$), and Mes ($O-SO_2-CH_3$),

Z is CH_2, $NH-C(O)-CH_2$, or $O-C(O)-CH_2$

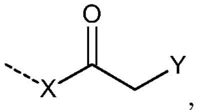
each dashed line in E indicating a site where E is bound to L for use in a method for increasing the stability of a protein, wherein the protein comprises at least 70 amino acids according to claim 1.

10. The trivalent thiol-reactive cross-linker of claim 9, wherein

L is 

,

U is CH_2, preferably n is 2 or 3; and

E is 

,

wherein X is NH and Y is F, Cl, or Br, preferably Cl.

11. Use of the cross-linker of claim 9 for reacting with thiol groups, preferably for cross-linking three cysteine resides present in a protein.

12. A protein comprising at least 70 amino acids and comprising at least three cysteine residues, wherein three of the cysteine residues each is covalently bonded to a trivalent thiol-reactive cross-linker, wherein the tertiary structure of the protein is stabilized as compared to the non-cross-linked protein.

13. The protein of claim 12, wherein the trivalent thiol-reactive cross-linker is a cross-linker according to claim 9.

14. The protein according to any of claims 11 and 12, wherein the protein is a Sortase A polypeptide comprising amino acid substitutions with cysteine at positions 111, 149, and 177, with reference to amino acid position numbering of Staphylococcus aureus SrtA, preferably wherein the polypeptide has SEQ ID NO: 1.

15. The protein according to any of claims 11 and 12, wherein the protein is a KIX domain polypeptide comprising amino acid substitutions with cysteine at positions 594, 599, and 646, with reference to amino acid position numbering of the following sequence GVRKG WHEHVTQDLR SHLVHKLVQA IFPTPDPAAL KDRRMENLVA YAKKVEGDMY ESANSRDEYY HLLAEKIYKI QKELEEKRRS R.

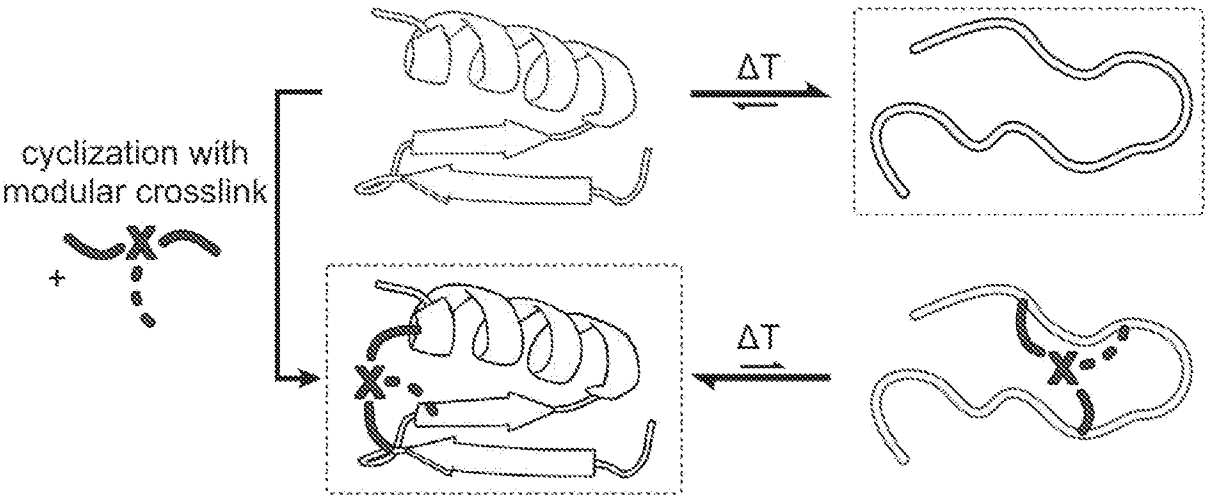

cyclization with
modular crosslink
+
ΔT
ΔT

Fig. 1a

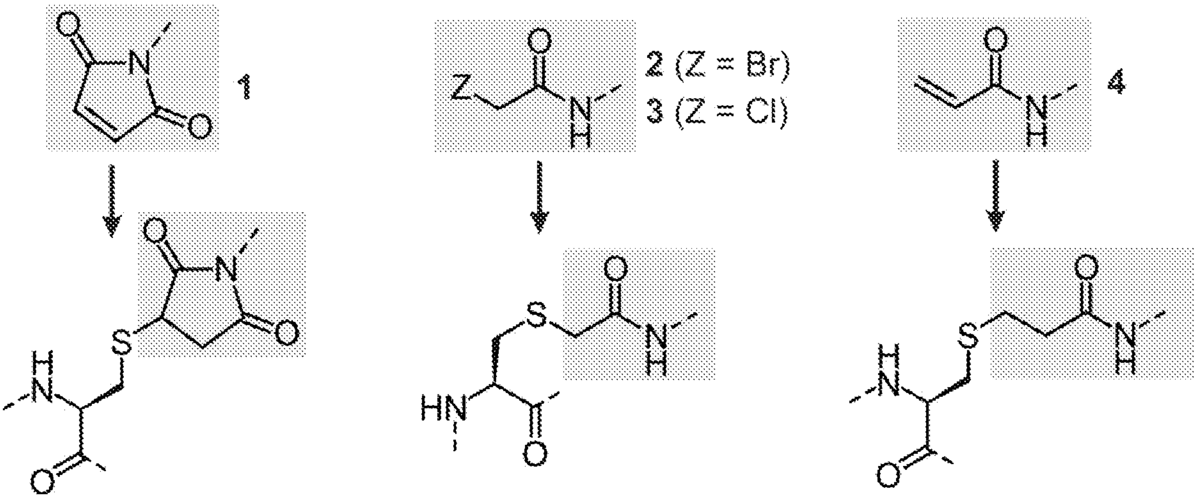

1
2 (Z = Br)
3 (Z = Cl)
4
Z
S
N
HN
O

Fig. 1b

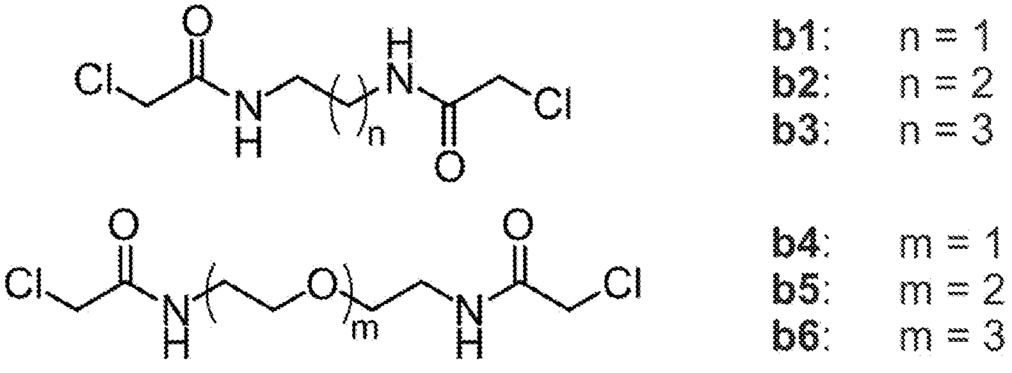

Cl
O
N
H
H
N
n
O
Cl
b1: n = 1
b2: n = 2
b3: n = 3
Cl
O
N
H
O
m
N
H
O
Cl
b4: m = 1
b5: m = 2
b6: m = 3

Fig. 2a

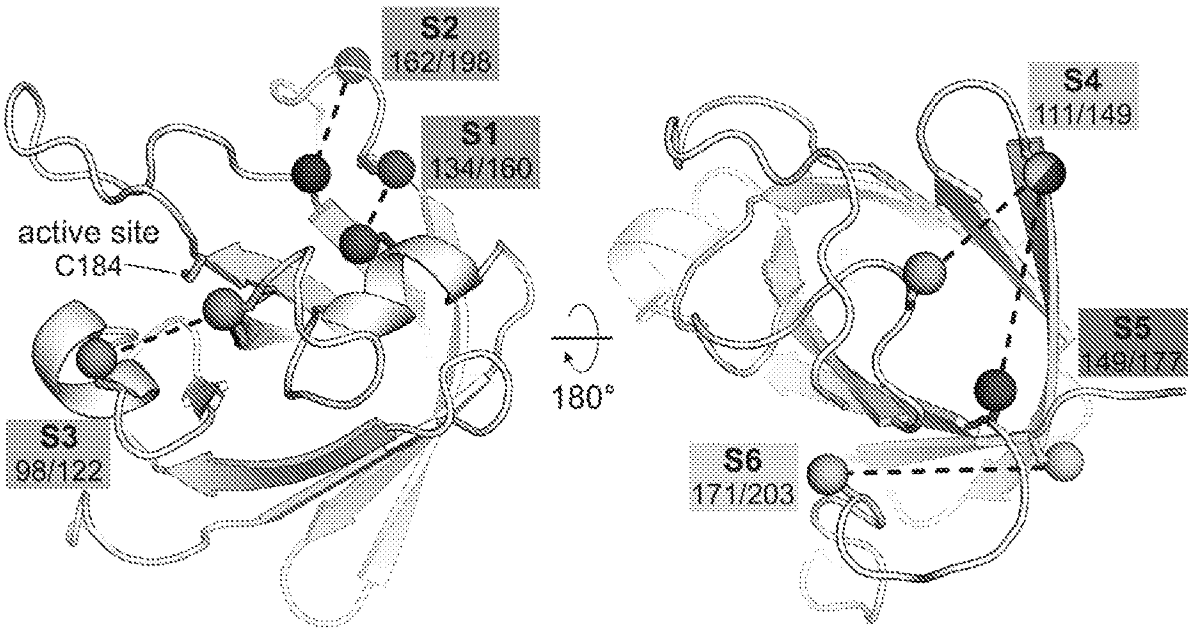

S2
162/198
S1
134/160
active site
C184
S3
98/122
180°
S4
111/149
S5
149/177
S6
171/203

Fig. 2b

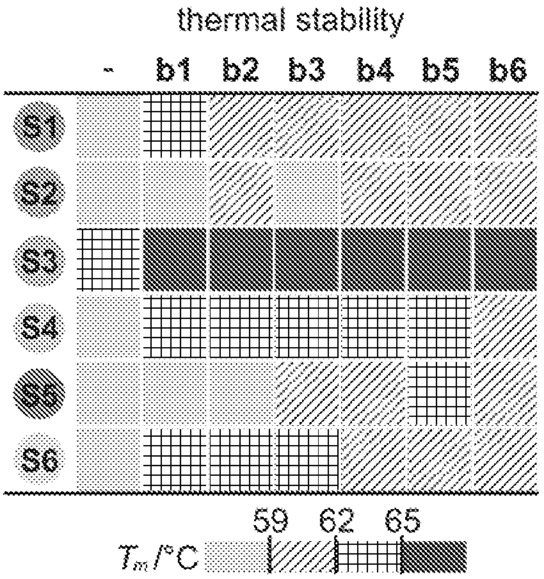


Fig. 2c

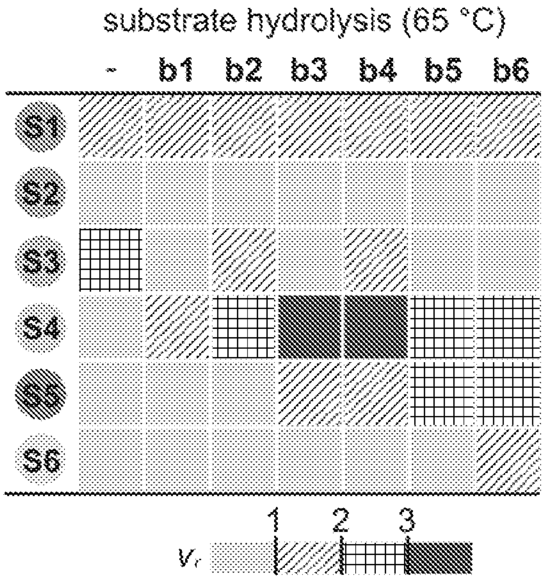


Fig. 2e

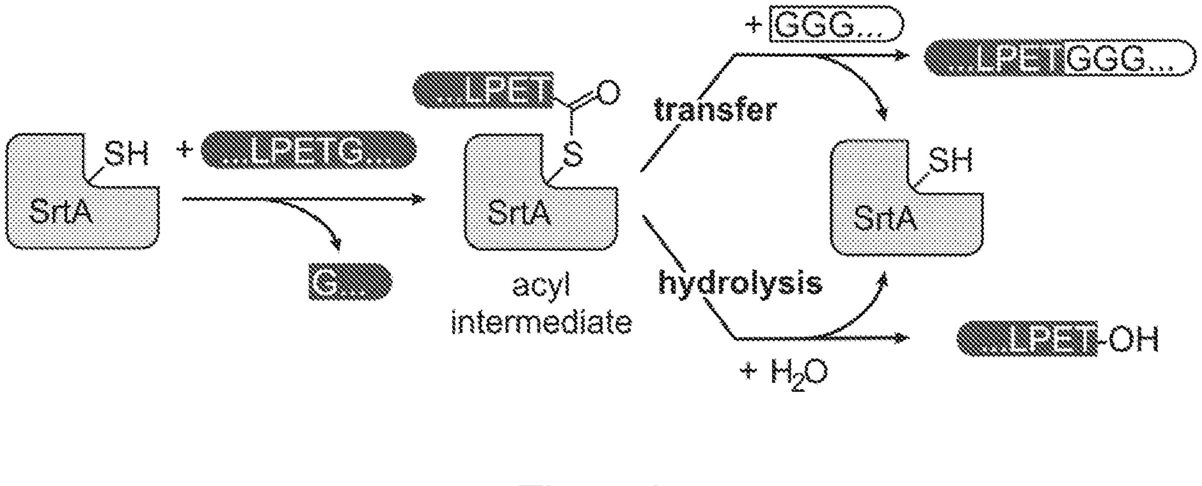


Fig. 2d

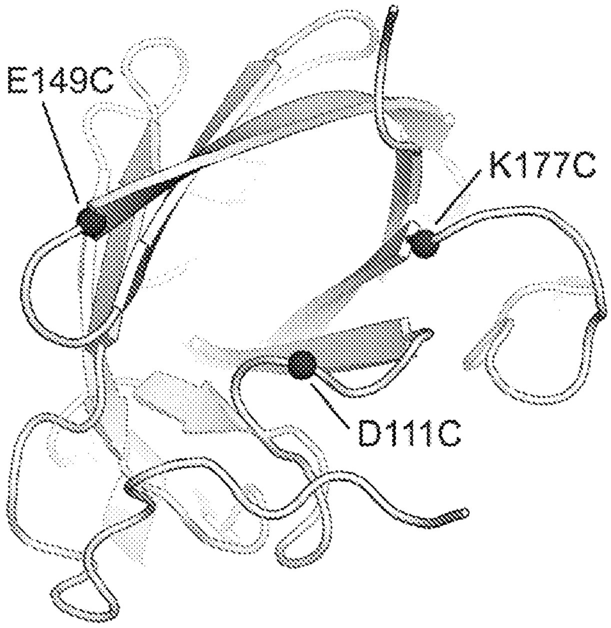

E149C
K177C
D111C

Fig. 3a

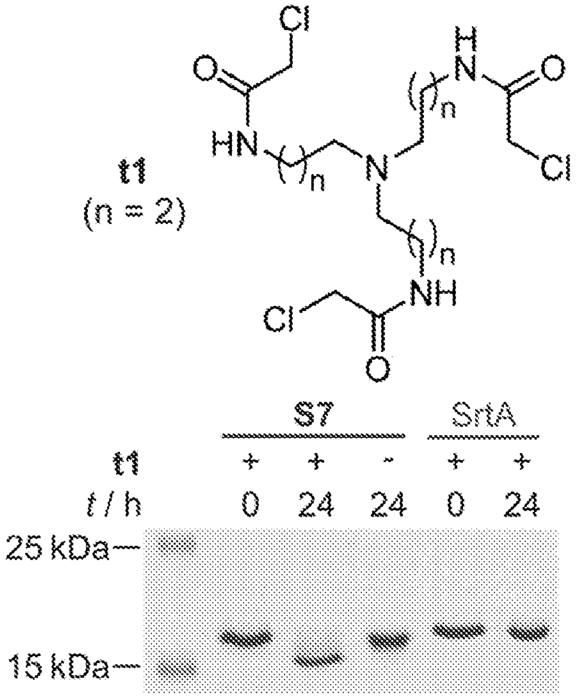

t1
(n = 2)
Cl
O
HN
n
N
H
NH
S7
SrtA
t1
+
+
-
+
+
t / h
0
24
24
0
24
25 kDa
15 kDa

Fig. 3b

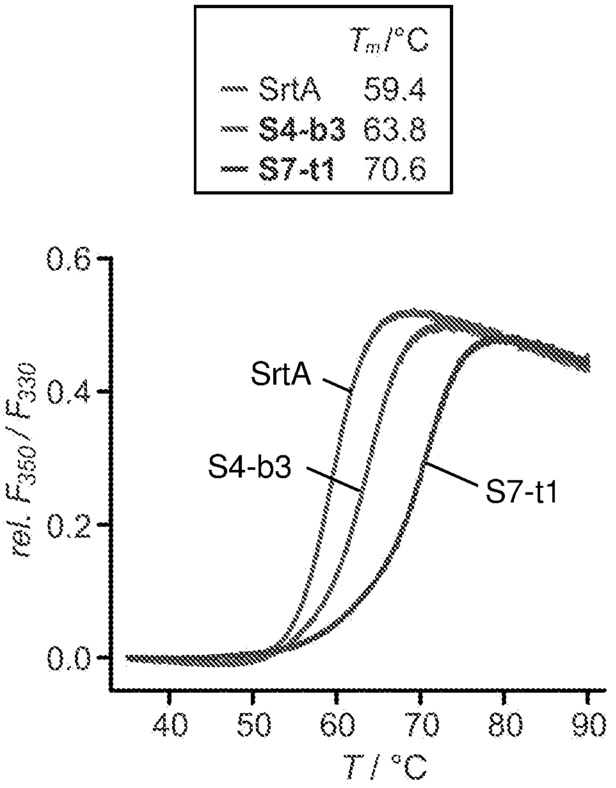

T_m/°C
SrtA 59.4
S4-b3 63.8
S7-t1 70.6
0.6
0.4
0.2
0.0
rel. F_{350} / F_{330}
SrtA
S4-b3
S7-t1
40
50
60
70
80
90
T / °C

Fig. 3c

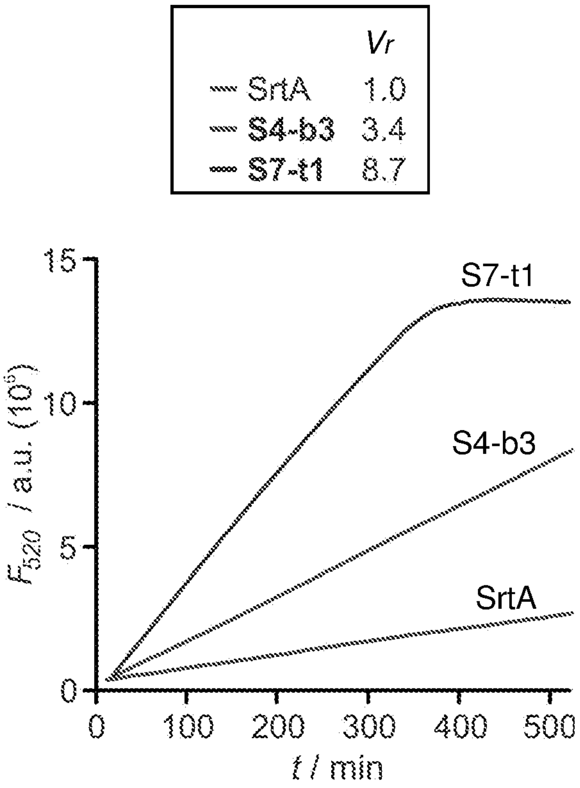

Vr
SrtA 1.0
S4-b3 3.4
S7-t1 8.7
15
10
5
0
F_{520} / a.u. (10^5)
S7-t1
S4-b3
SrtA
0
100
200
300
400
500
t / min

Fig. 3d

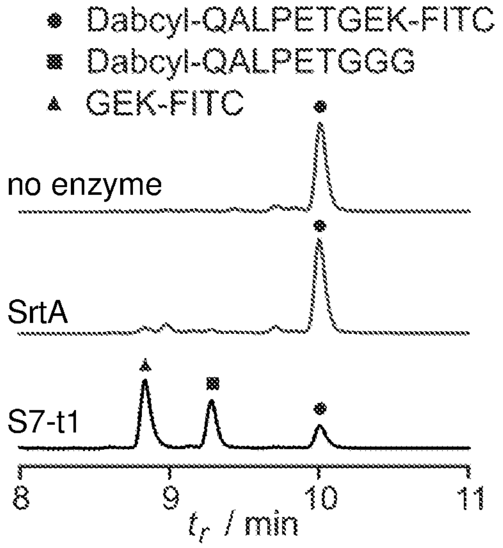


Fig. 4a

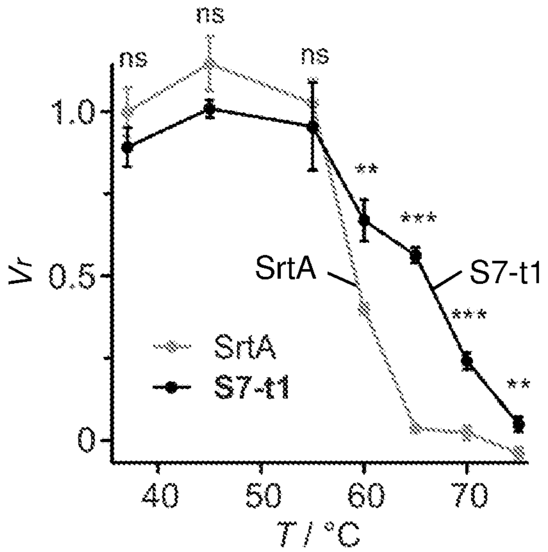


Fig. 4b

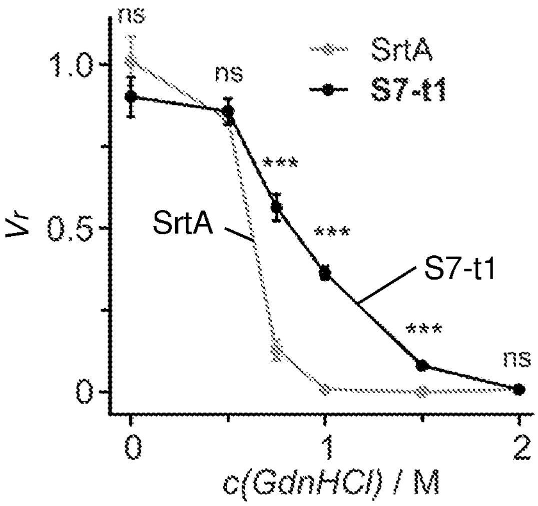


Fig. 4c

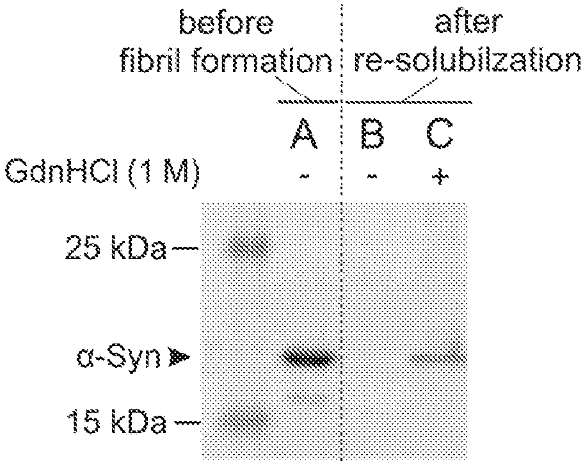


Fig. 4d

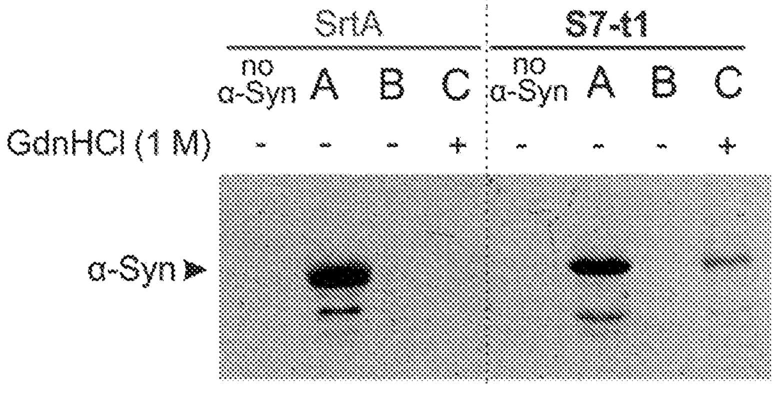
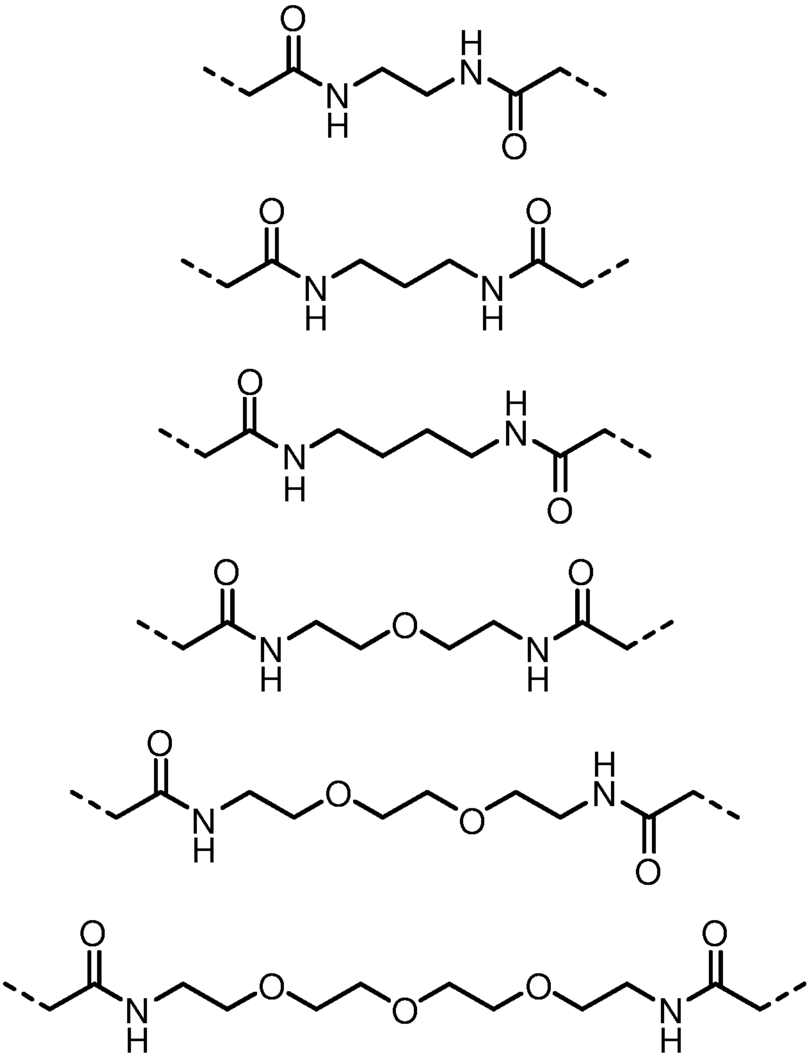


Fig. 4e



b1	11.0 Å
b2	12.3 Å
b3	13.5 Å
b4	14.5 Å
b5	18.0 Å
b6	21.4 Å

Fig. 5

60 70 80 90 100
GSHMQ AKPQIPKDKS KVAGYIEIPD ADIKEPVYPG PATPEQL**N**RG

110 120 130 140 150
VSFAEENESL **D**DQNISIAGH T**F**IDRPNYQF TNL**K**AAKKGS MVYFKVGN**E**T

160 170 180 190 200
RKYKMTSIR**D** V**K**PTDVGVLD **E**QKGKD**K**QLT LITCDDYNEK TGVWEKR**K**IF

VA**T**EVK

Fig. 6a

Variant	Mutation	Distance (C_α- C_α)
S1	K134C-D160C	9.0 Å
S2	K162C-K198C	8.4 Å
S3	N98C-F122C	11.2 Å
S4	D111C-E149C	12.2 Å
S5	E149C-K177C	16.5 Å
S6	E171C-T203C	11.8 Å

Fig. 6b

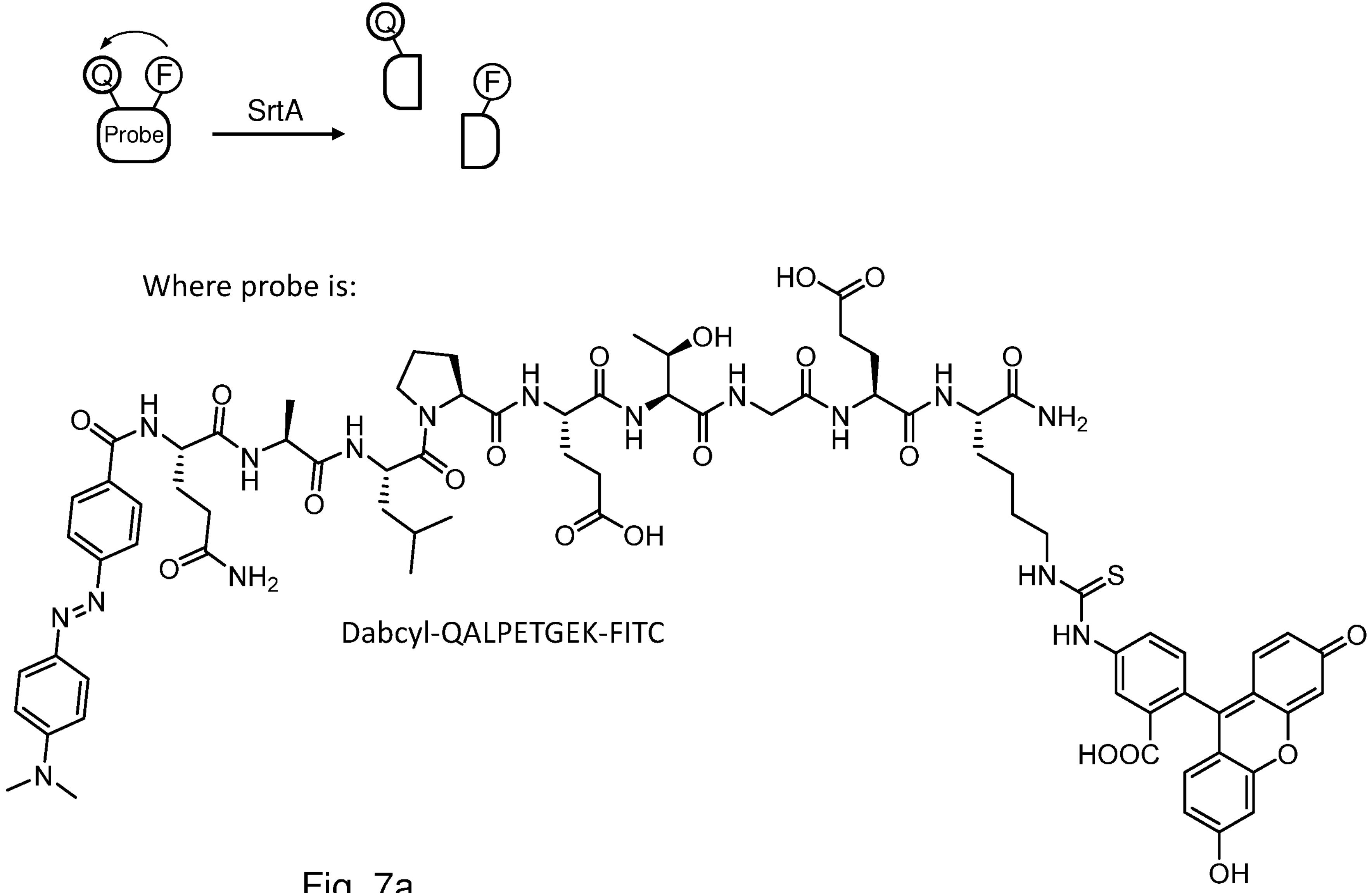


Fig. 7a

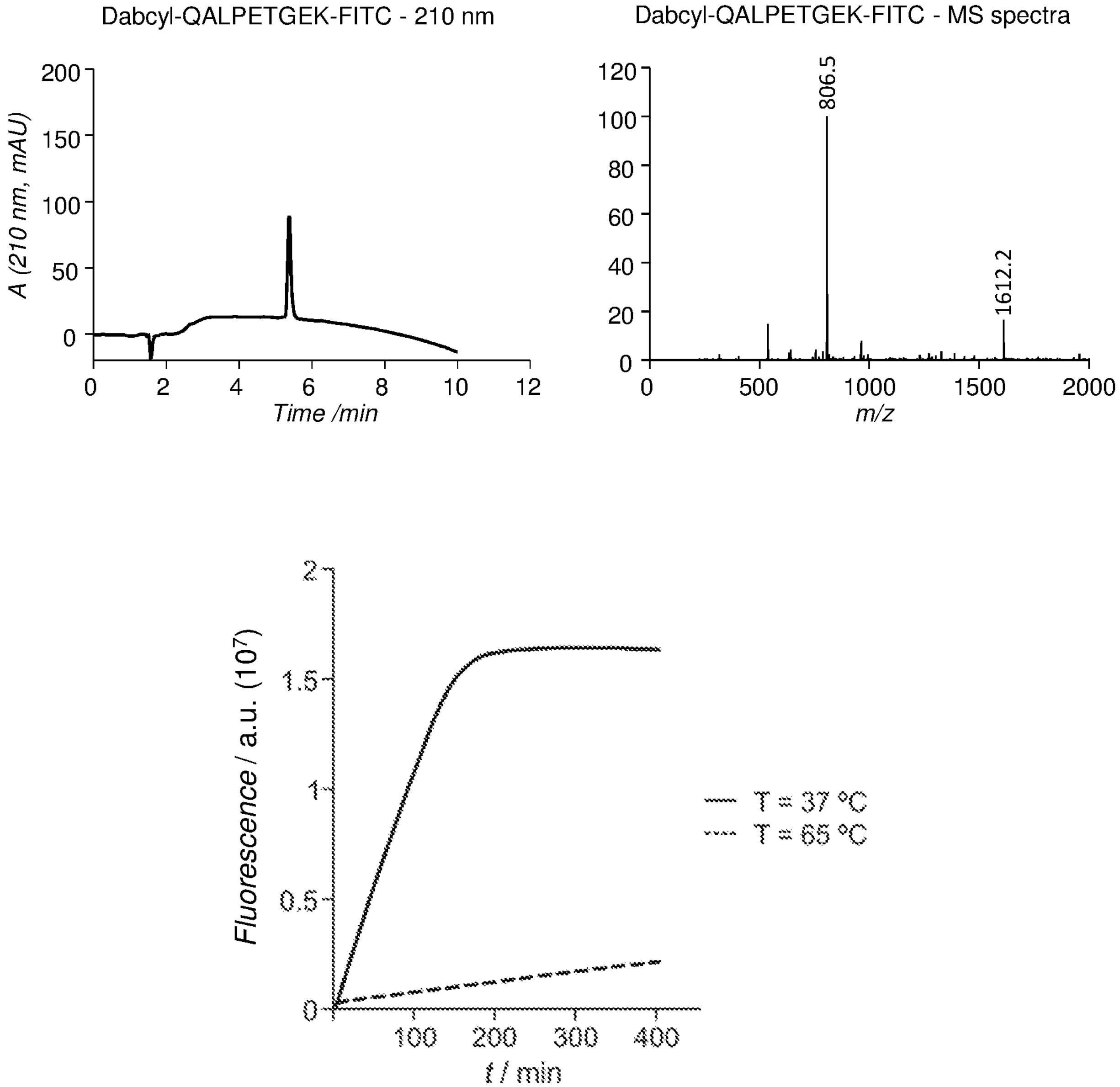
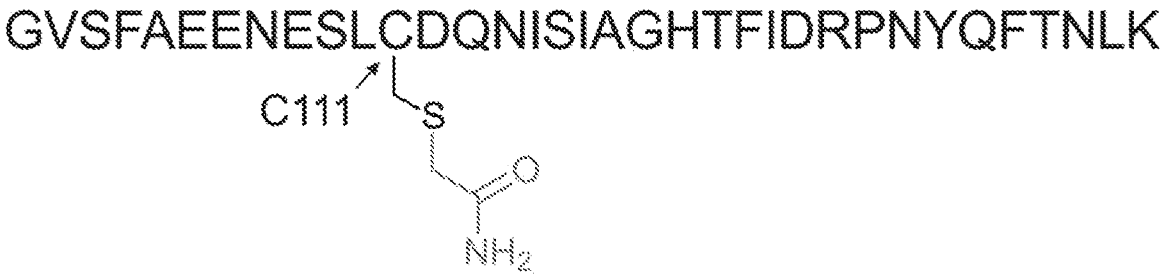

Dabcyl-QALPETGEK-FITC - 210 nm
A (210 nm, mAU)
200
150
100
50
0
0
2
4
6
8
10
12
Time /min
Dabcyl-QALPETGEK-FITC - MS spectra
806.5
1612.2
120
100
80
60
40
20
0
0
500
1000
1500
2000
m/z
Fluorescence / a.u. (10^7)
2
1.5
1
0.5
0
100
200
300
400
t / min
T = 37 °C
T = 65 °C

Fig. 7b



molecular formula (monoisotopic m):
$C_{173}H_{261}N_{47}O_{58}S$ (4013.89 g/mol)

calc. monoisotopic peak for MH^{3+}:
m/z = 1338.97 th

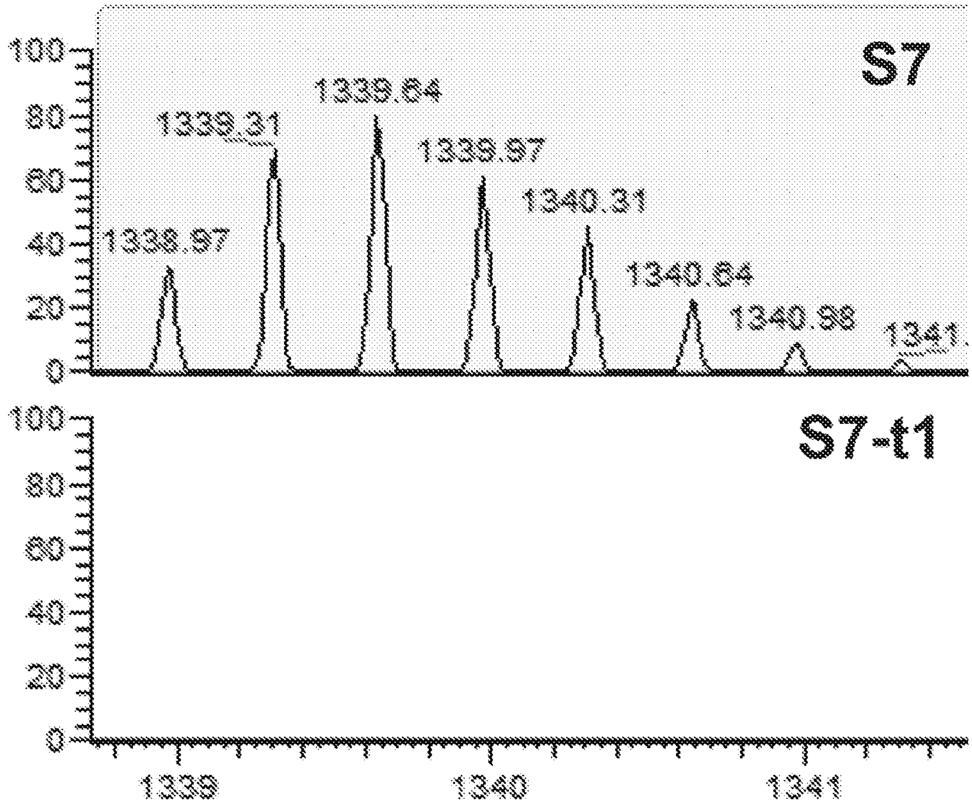
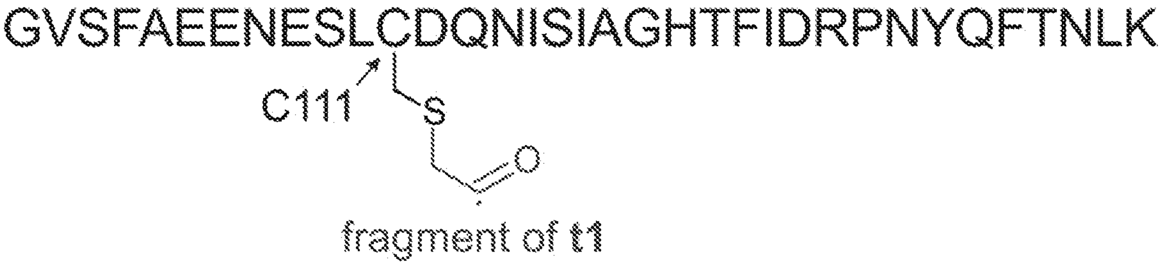


Fig. 8a



molecular formula (monoisotopic m):
$C_{175}\,H_{262}\,N_{47}\,O_{59}\,S$ (3997.87 g/mol)

calc. monoisotopic peak for MH^{3+}:
m/z = 1333.63 th

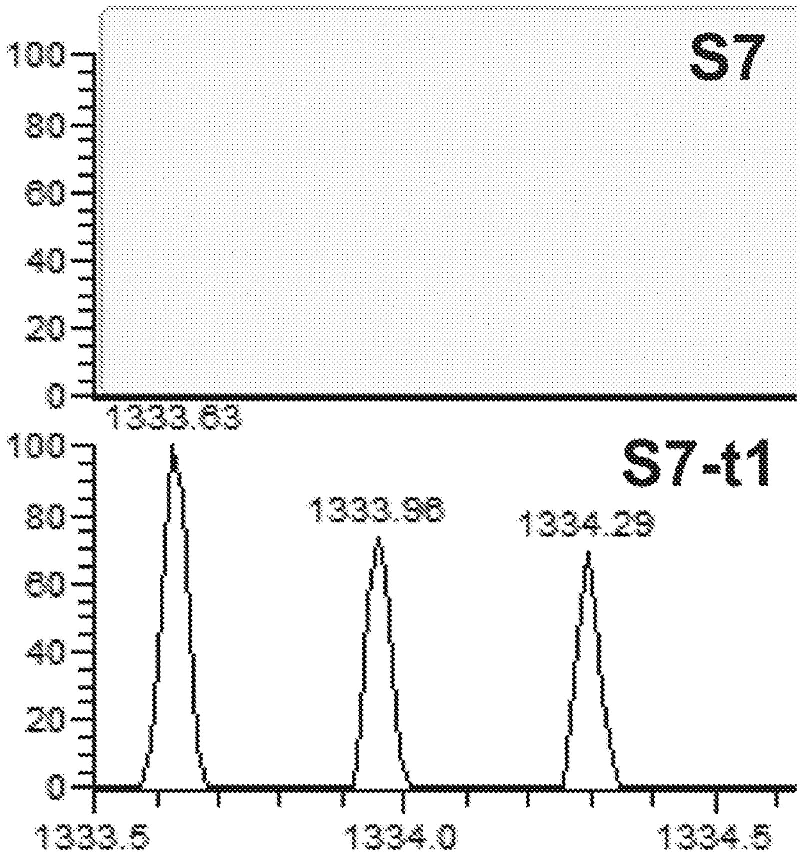


Fig. 8a, Cont'd

molecular formula (monoisotopic m):
$C_{69} H_{112} N_{20} O_{19} S_2$ (1588.79 g/mol)

calc. monoisotopic peak for MH^{3+}:
m/z = 530.60 th

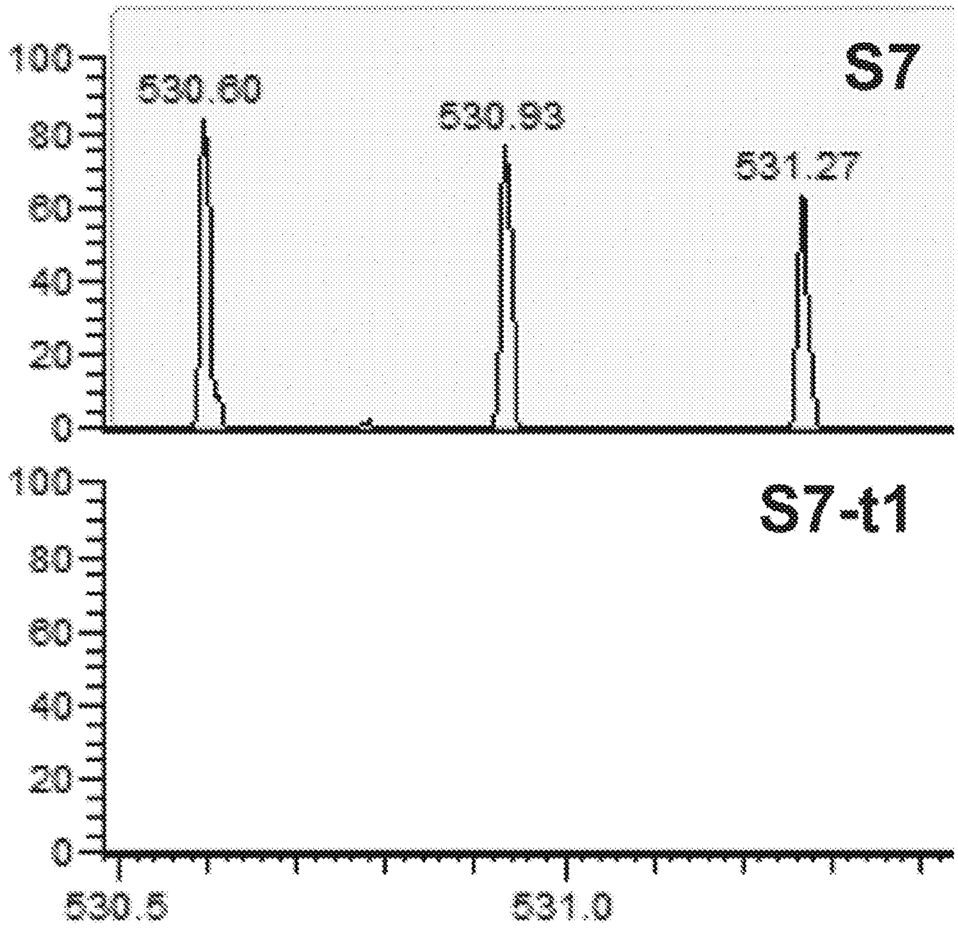
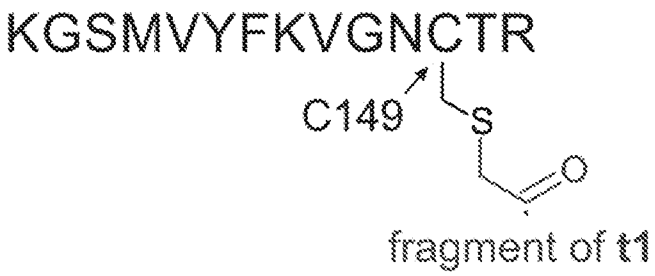


Fig. 8b



molecular formula (monoisotopic m):
$C_{71}\ H_{113}\ N_{20}\ O_{20}\ S_{2}$ (1629.79 g/mol)

calc. monoisotopic peak for MH^{2+}:
m/z = 815.90 th

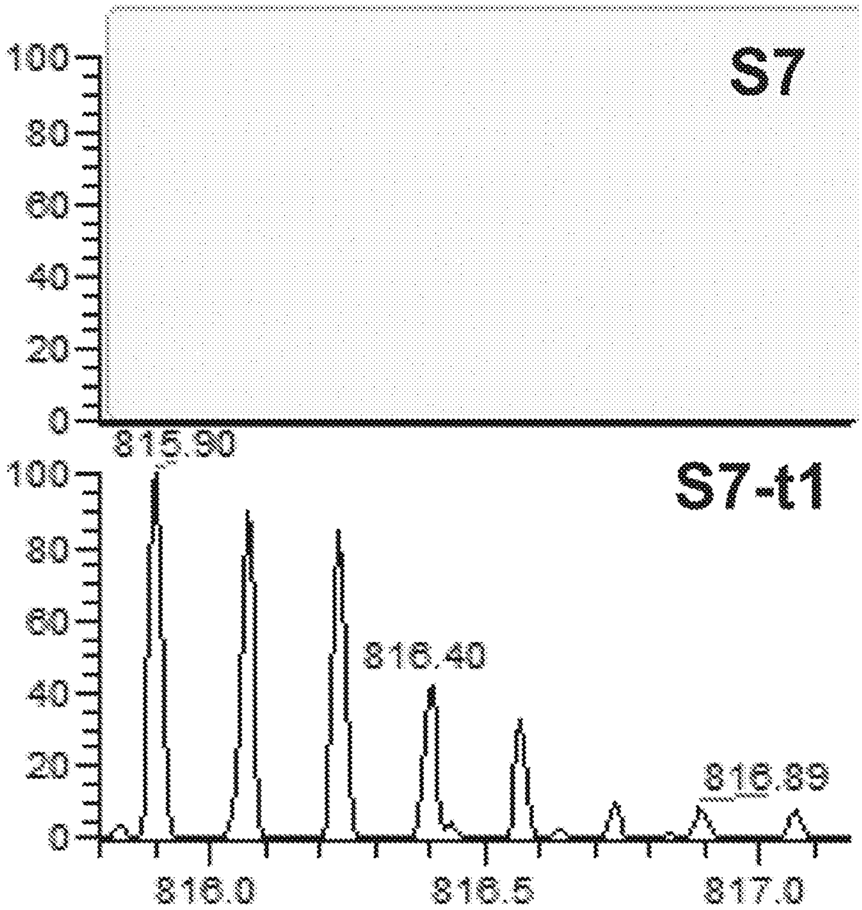
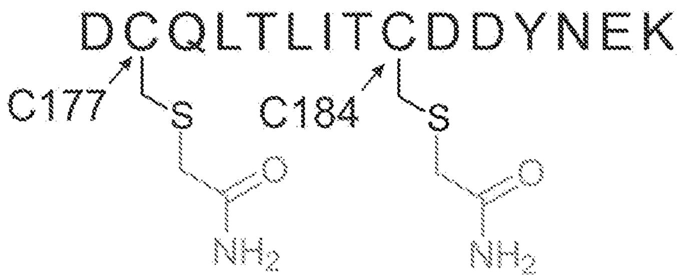


Fig. 8b, Cont'd



molecular formula (monoisotopic m):
$C_{77}\,H_{122}\,N_{20}\,O_{31}\,S_2$ (1886.80 g/mol)

calc. monoisotopic peak for MH^{3+}:
m/z = 629.94 th

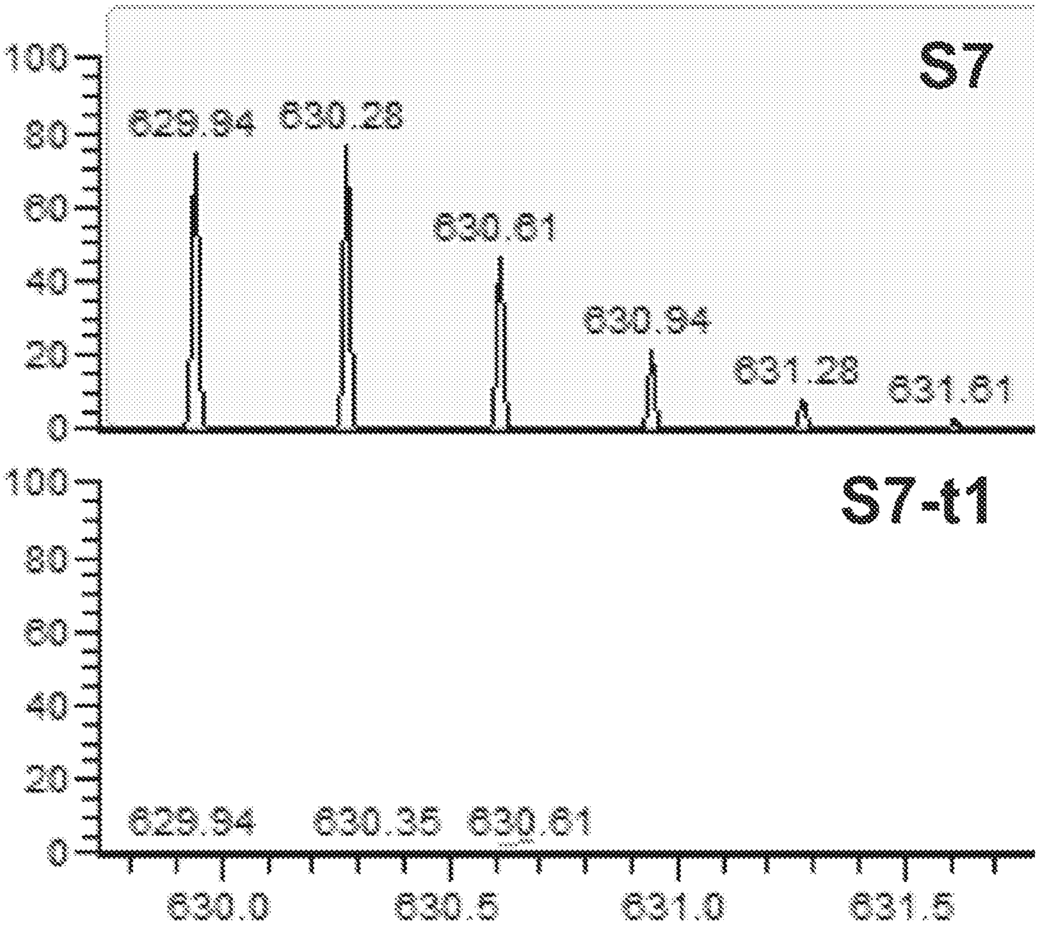
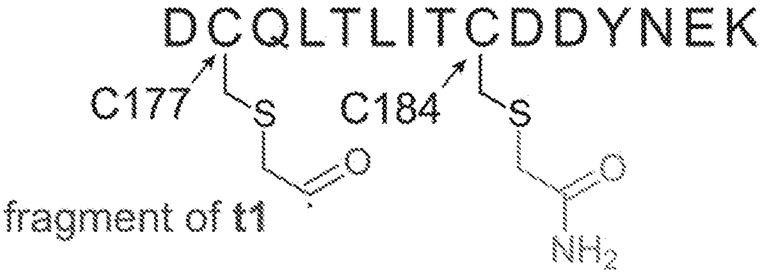


Fig. 8c



molecular formula (monoisotopic m):
$C_{77}H_{119}N_{18}O_{32}S_2$ (1871.77 g/mol)

calc. monoisotopic peak for MH^{3+}:
m/z = 936.89 th

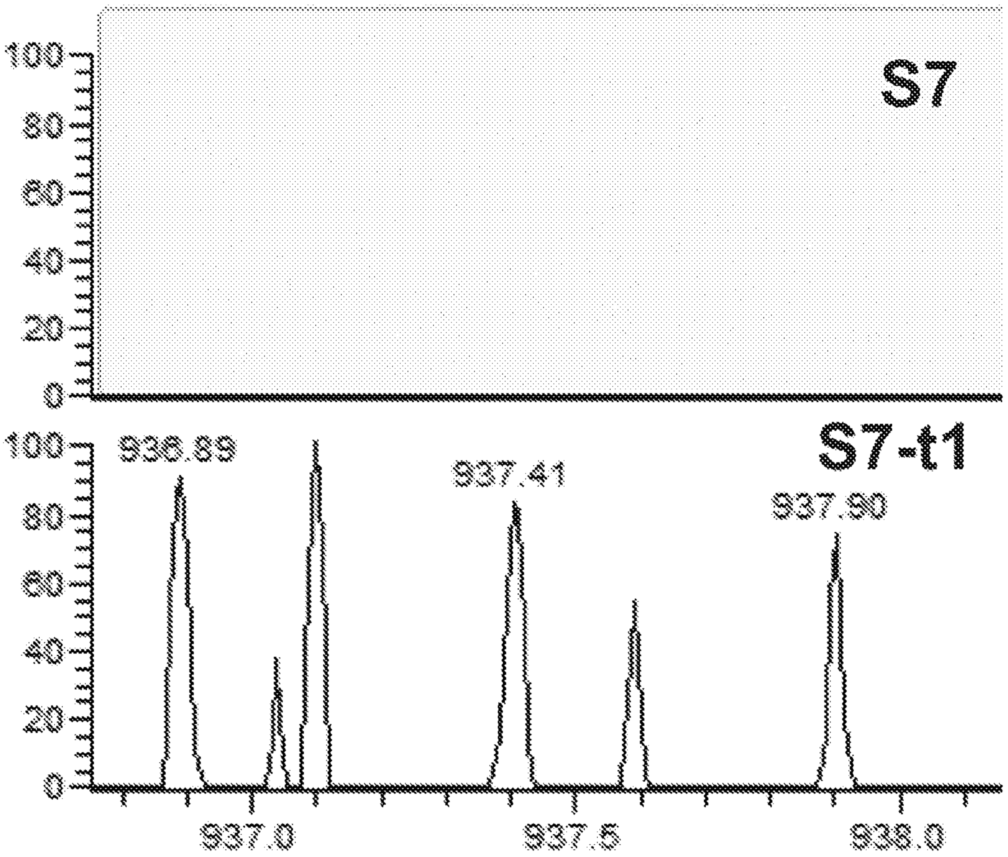


Fig. 8c, Cont’d

```
        10         20         30         40         50
MDVFMKGLSK AKEGVVAAAE KTKQGVAEAA GKTKEGVLYV GSKTKEGVVH
        60         70         80         90        100
GVATVAEKTK EQVTNVGGAV VTGVTAVAQK TVEGAGSIAA ATGFVKKDQL
       110        120        130        140        150
GKNEEGAPQE GILEDMPVDP DNEAYEMPSE EGYQDYEPEA GGGGSLPETG

HHHHHH  (SEQ ID NO:2
```

Fig. 9a

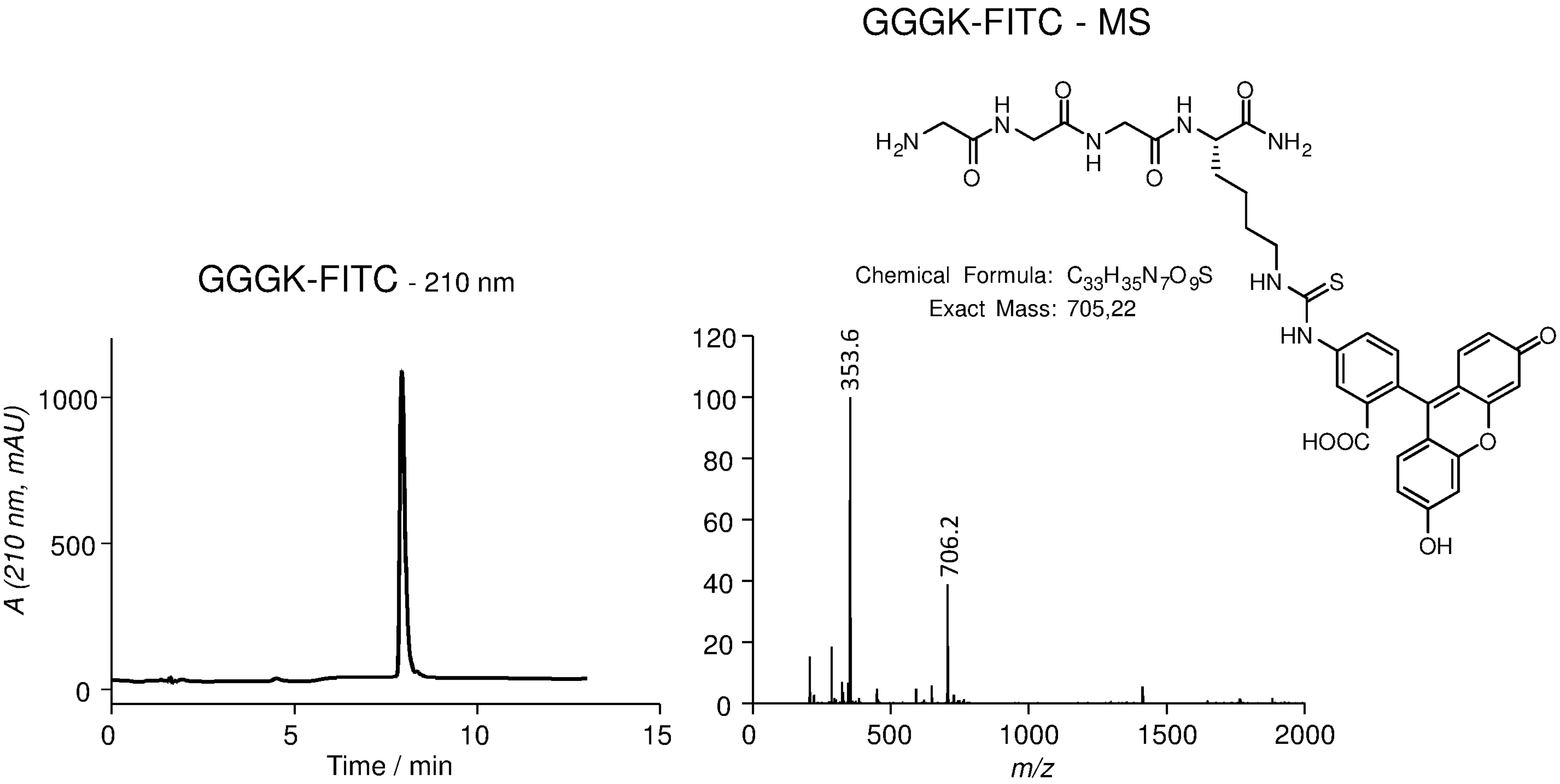

GGGK-FITC - 210 nm
A (210 nm, mAU)
1000
500
0
0
5
10
15
Time / min
GGGK-FITC - MS
H₂N
H
N
O
N
H
NH₂
HN
S
HN
HOOC
OH
Chemical Formula: C₃₃H₃₅N₇O₉S
Exact Mass: 705,22
120
100
80
60
40
20
0
353.6
706.2
0
500
1000
1500
2000
m/z

Fig. 9b

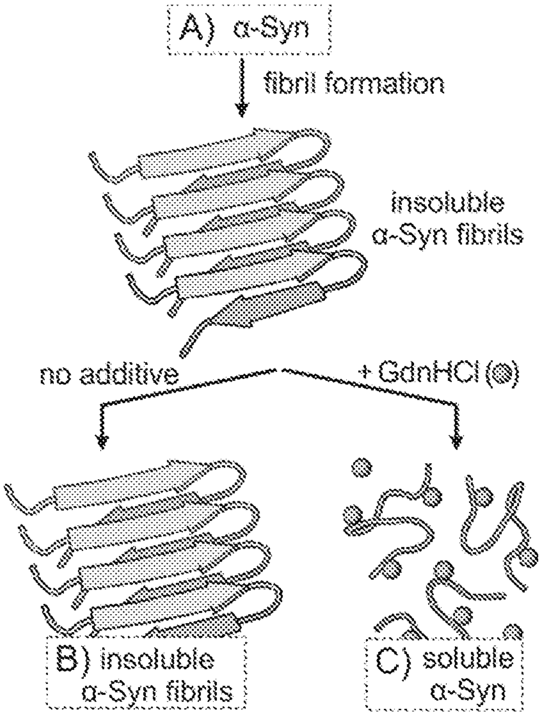


Fig. 10a

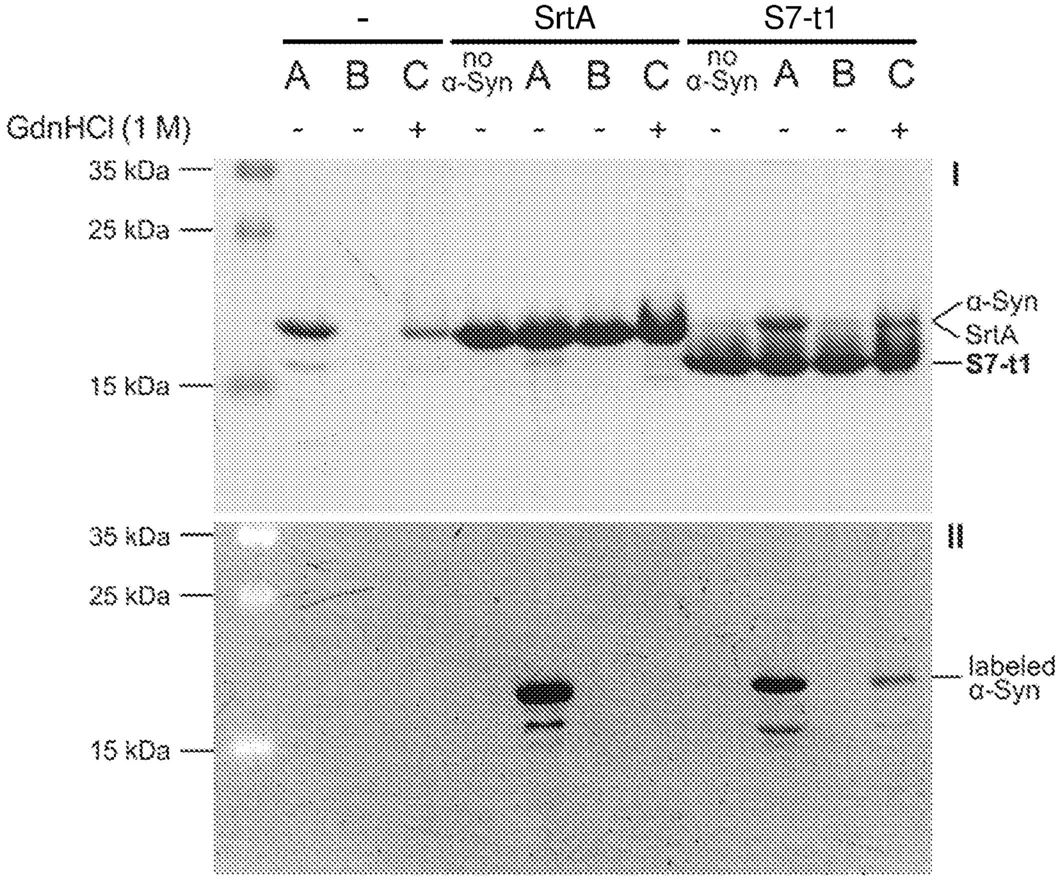


Fig. 10b

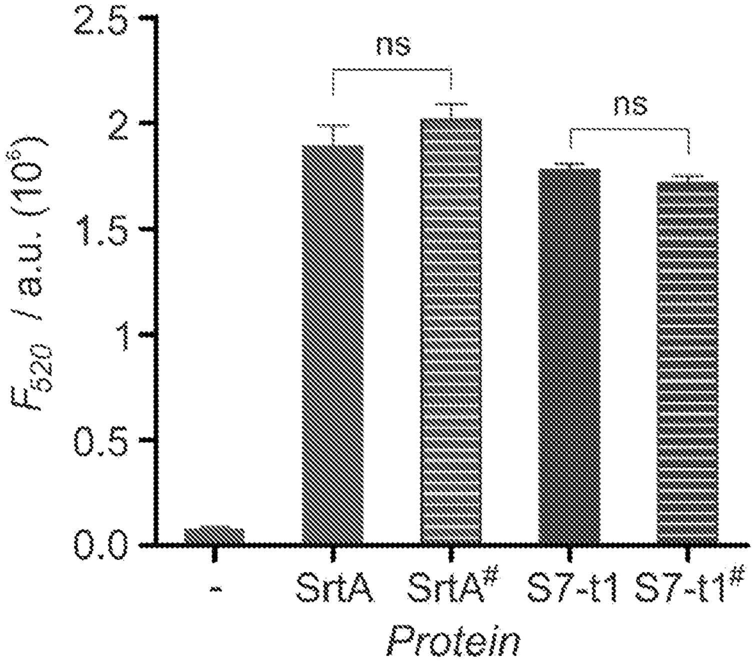


Fig. 11

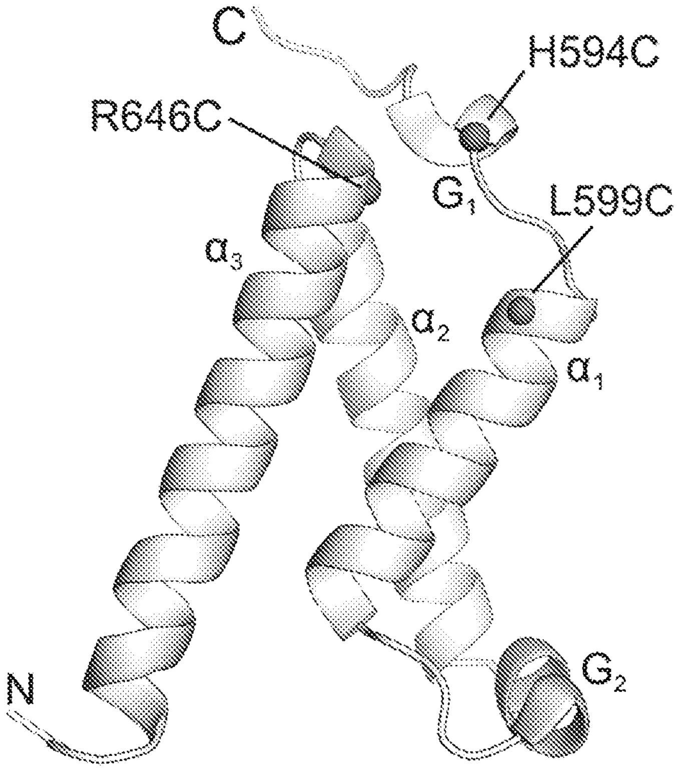


Fig. 12a

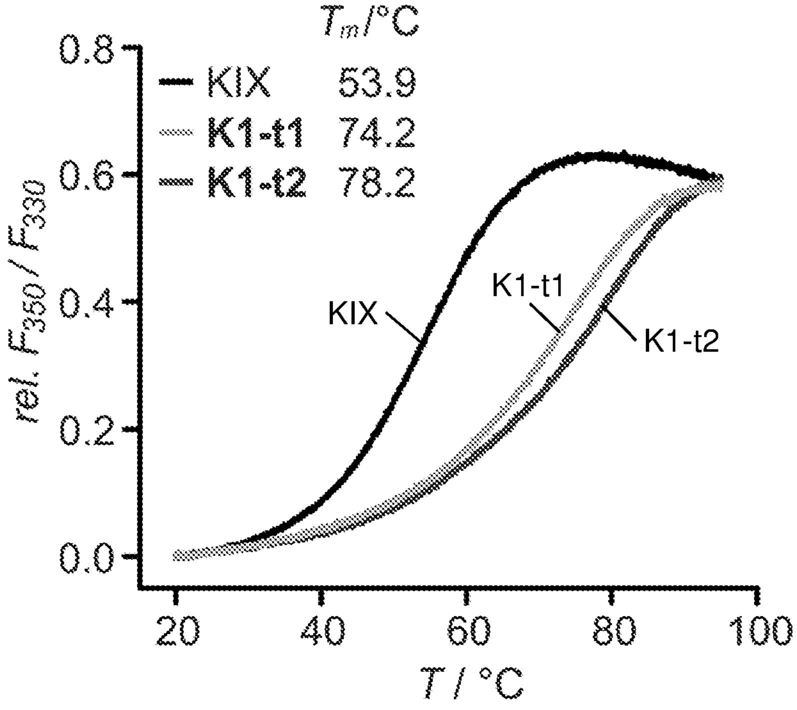


Fig. 12b

KIX wt:

```
               590        600        610        620        630
          GVRKG WHEHVTQDLR SHLVHKLVQA IFPTPDPAAL KDRRMENLVA

        640        650        660        670
YAKKVEGDMY ESANSRDEYY HLLAEKIYKI QKELEEKRRS R
```

Fig. 13a

K1:

```
               590        600        610        620        630
          GVRKG WHEHVTQDLR SHLVHKLVQA IFPTPDPAAL KDRRMENLVA

        640        650        660        670
YAKKVEGDMY ESANSRDEYY HLLAEKIYKI QKELEEKRRS R
```

Fig. 13b

Variant	Mutation	Distance (C_α- C_α)
K1	H594C-L599C-R646C	10.0 Å
	H594C-L599C-R646C	11.5 Å
	H594C-L599C-R646C	7.8 Å

Fig. 13c

Variant	Mutation	Distance (C_α- C_α)
S7	D111C-E149C-K177C	12.4 Å
	D111C-E149C-K177C	15.7 Å
	D111C-E149C-K177C	8.5 Å

Fig. 13d

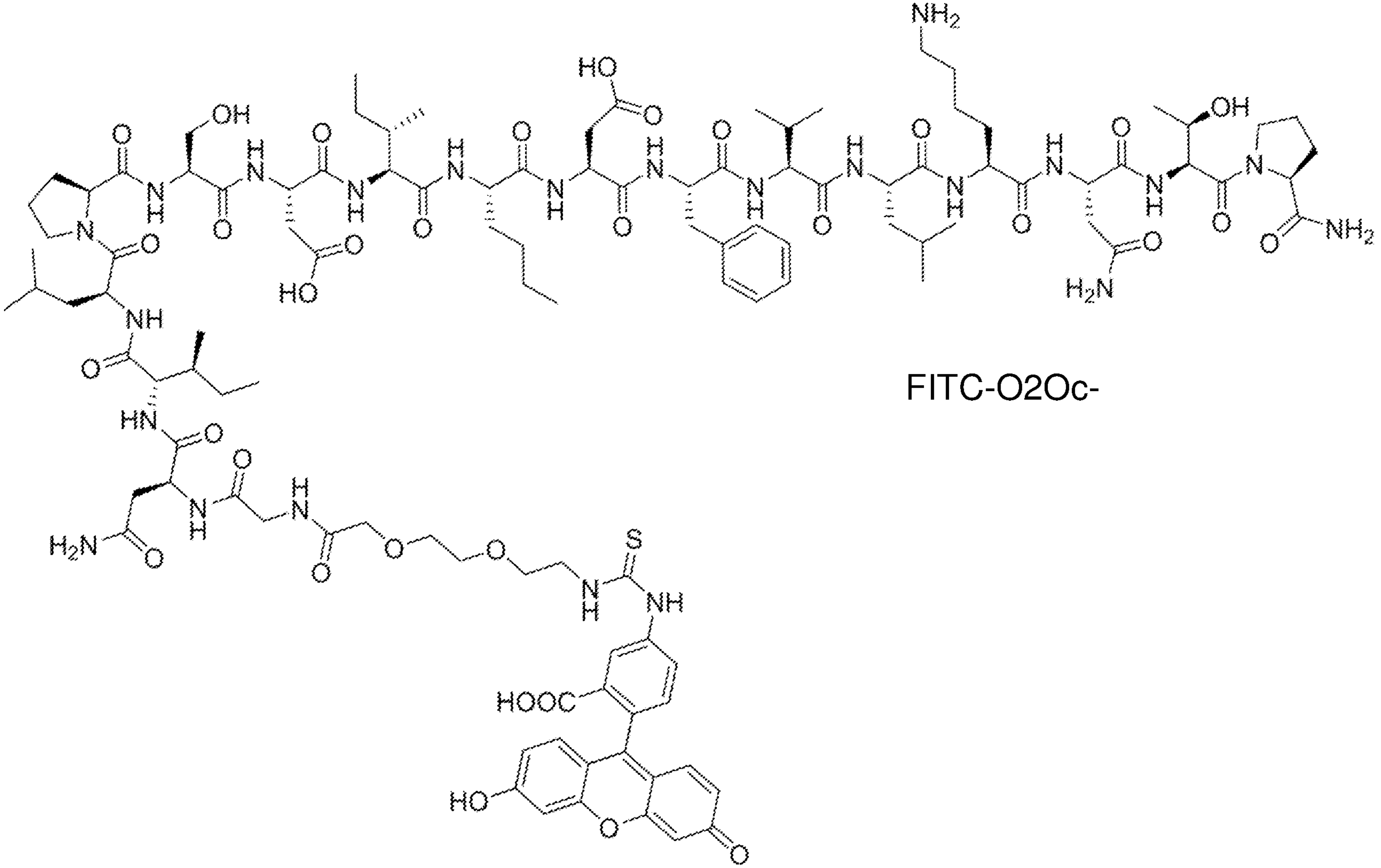


Fig. 14a

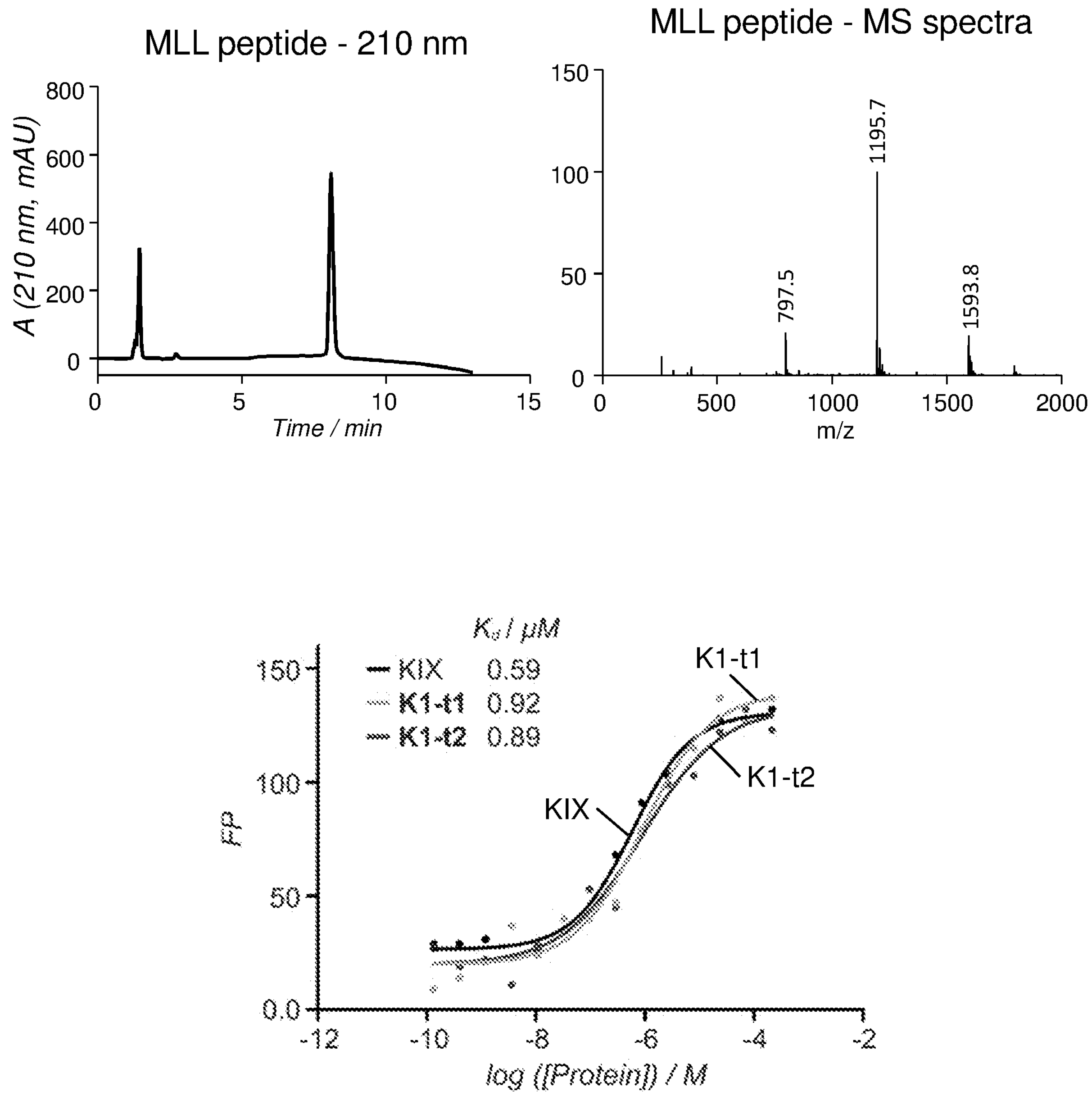

MLL peptide - 210 nm
A (210 nm, mAU)
800
600
400
200
0
0
5
10
15
Time / min
MLL peptide - MS spectra
150
100
50
0
1195.7
797.5
1593.8
0
500
1000
1500
2000
m/z
K_d / µM
KIX 0.59
K1-t1 0.92
K1-t2 0.89
K1-t1
K1-t2
KIX
FP
150
100
50
0.0
-12
-10
-8
-6
-4
-2
log ([Protein]) / M

Fig. 14b

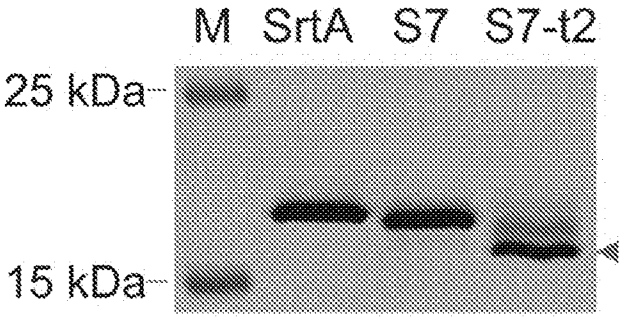


Fig. 15a

Ions	S7-t2 m/z calc.	m/z found
$[M+1H]^{+}$	17212.6	-
$[M+9H]^{9+}$	1913.5	1913.3
$[M+10H]^{10+}$	1722.3	1722.2
$[M+11H]^{11+}$	1565.8	1565.5
$[M+12H]^{12+}$	1435.4	1435.1
$[M+13H]^{13+}$	1325.0	1324.8
$[M+14H]^{14+}$	1230.5	1230.6
$[M+14H]^{14+}$	1148.5	1148.5
$[M+16H]^{16+}$	1076.8	1076.5
$[M+17H]^{17+}$	1013.5	1013.3
$[M+18H]^{18+}$	957.3	957.0

Fig. 15b

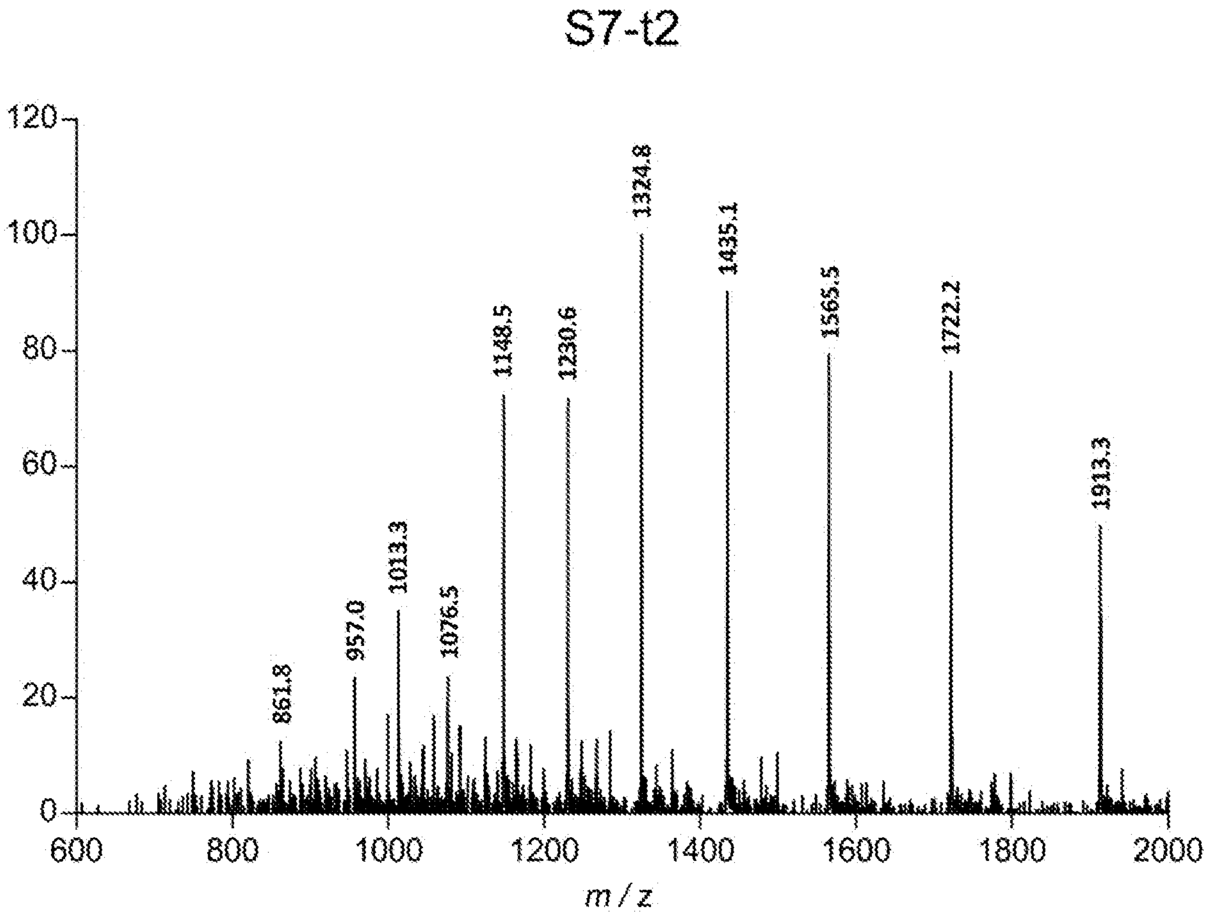


Fig. 15c

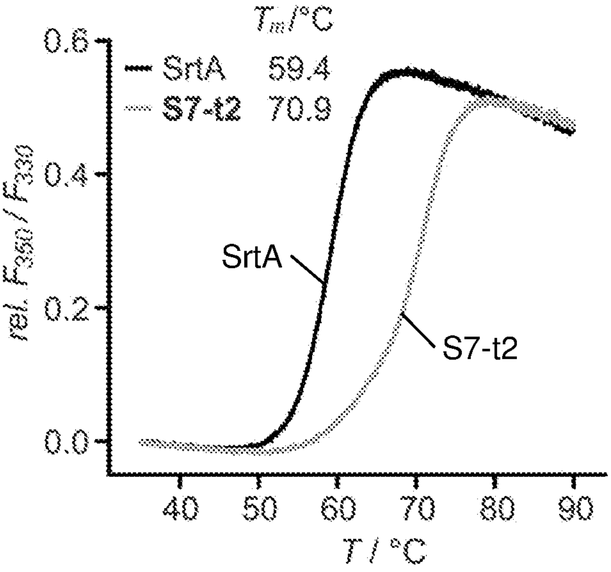


Fig. 15d

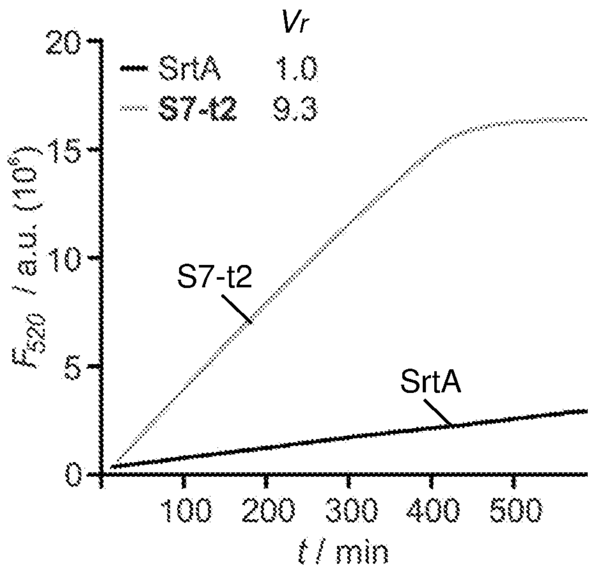


Fig. 15e