Supplemental Data

A HIGHLY UNUSUAL THIOESTER BOND IN A PILUS ADHESIN IS REQUIRED FOR EFFICIENT HOST CELL INTERACTION

Pointon JA[1], Smith WD[1], Saalbach G[2], Crow A[2], Kehoe MA[1], & Banfield MJ[2]
[1]Institute for Cell and Molecular Biosciences, Newcastle University, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK, [2]Dept. of Biological Chemistry, John Innes Centre, Colney Lane, Norwich, NR4 7UH, UK

Supplemental Text

Sequence conservation of Spy0125 in GAS

Sequence alignments of Spy0125 homologues are shown in Supplemental Figure 1A (CTR) and Supplemental Figure 1B (NTR). Overall, the sequences of these proteins are well conserved across GAS serotypes. Key residues that comprise the thioester and isopeptide bond linkages, and their immediate surrounds are conserved. The lack of an N-terminal region in the M5 and M18 strains of GAS is worth noting. Also, despite being dispensable for Spy0125 binding to HaCaT cells, the NTR contains some sequence conservation, and a very similar spatial distribution, at the residue positions that comprise the thioester in the CTR. There is, however, only limited sequence identity (~17 %) between the Spy0125-NTR and the region that comprises the top domain of Spy0125-CTR. Alignments were produced using CLUSTALW2 (1).

Additional description of structures

Within each of the models, the domains adopt a very similar structure to each other. Overlays of the middle domains with their equivalent regions in all structures generates rmsds over the range 0.25 – 0.50 Å (81 – 85 equivalent C_α positions); for the top domains a range of 0.06 – 0.55 Å is observed (185 – 194 equivalent C_α positions); and for the bottom domains the values are 0.053 – 0.33 Å (115 – 117 equivalent C_α positions, note: the lower values of the range represent rmsds allowed by the NCS restraints, which were maintained in the last rounds of refinement).

There is a single region in the structures of Spy0125-CTR that adopts a different conformation in the A-form and B-form crystals (not counting regions discussed elsewhere as prone to rearrangements/disorder). In each crystal form the region including residues Asp439 – Val453 is well defined in the electron density. Crystal contacts within this region differ between the structures and this is the most likely reason for this subtle rearrangement.

Disorder in the Spy0125-CTR crystals

Analysis of the fit of the models to the electron density in each of the Spy0125-CTR structures reveals some degree of structural disorder/flexibility within the protein at the domain level. Residual electron density within the A-form crystals, adjacent to the modeled C-terminus of each

chain, suggests the presence of the bottom domain in the crystal. However, it is not possible to reliably position this region, even though there is space in the lattice (Supplemental Figure 2A). Molecular replacement calculations using the final model of the bottom domain from the B-form crystals reveals positions for this domain allowable by packing constraints in the A-form lattice. Despite this, subsequent refinement in REFMAC5 (with these domains included) reveals an essentially unchanged R_{free} and inspection of resulting electron density maps does not allow an atomic model for this domain to be constructed. Therefore, this domain is not included in the final model, which leads to an apparent non-continuous crystal lattice (Supplemental Figure 2A). This kind of crystallographic disorder is not without precedent (2).

In the model derived from the B-form crystals, with data collected at the ESRF to 2.9 Å, the region comprising residues B470 – B560 shows a significantly poorer fit to the electron density compared to the rest of the model and has higher than average atomic displacement parameters. This region comprises part of the top domain (Supplemental Fig. 2B) and refines in the same conformation as found in the two molecules of the A-form crystals, the B-form model based on the DLS data and the equivalent region in the A-chain of the same asymmetric unit. Removal of this region leads to a poorer model, as judged by an increase in R_{free}, therefore it is retained in the final structure.

In the B-form crystal structure, refined against the DLS data, the C-terminal (bottom) domain of the A-chain (residues 604 – 719) displays a rather poor fit to the electron density in a number of regions and displays higher than average atomic displacement parameters compared to the same region in the other structures determined (Supplemental Figure 2B). As refined, this region assumes the same conformation as that observed in the B-form crystals derived from the ESRF data and the B-chain of the same asymmetric unit, for which the electron density is considerably better. Removal of this region leads to a poorer model, as judged by an increase in R_{free}, therefore it is retained in the final structure.

REFERENCES

1. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) *Bioinformatics* **23**, 2947-2948
2. van Raaij, M. J., Schoehn, G., Burda, M. R., and Miller, S. (2001) *J Mol Biol* **314**, 1137-1146

Supplemental Figure Legends

Supplemental Fig. 1 Sequence alignment of Spy0125 (M1) homologues in Group A *Streptococci*, encompassing **(A)** Spy0125-CTR, **(B)** Spy0125-NTR. In **(A)**, residues to the C-terminus of the sortase signal sequence (not present in the mature protein) are not shown. To the N-terminus, the residues in gray font show the end of the N-terminal region (M1, M28, M12, M49, M3) or the predicted start of the mature protein (after signal peptide cleavage, M5 and M18); note the poly-proline signature in all homologues that have an N-terminal region. Residues that comprise the thioester and isopeptide bonds are highlighted in red and yellow/green respectively. The glutamates that catalyse formation of the isopeptides are shown in magenta, with the surrounding hydrophobic side chains in cyan. In **(B)**, residues in pink are those predicted to encompass the secretion signal sequence. Residues in gray are those in the CTR. Residues that have sequence homology and relative spatial separation to residues comprising the thioester in the CTR are in highlighted in yellow.

Supplemental Fig. 2 Disorder in the structures of Spy0125. (**A**) Approximately orthogonal views of the apparent non-continuous lattice in the A-form crystal structure of Spy0125-CTR when a model for the bottom domain is not included (the two molecules of one asymmetric unit are shown in red/green). Closest atom-atom distance ~ 16 Å between the layers in the plane of the image. **(B)** Regions of domain flexibility observed in certain molecules in the structures of Spy0125-CTR (see Supplemental Text).

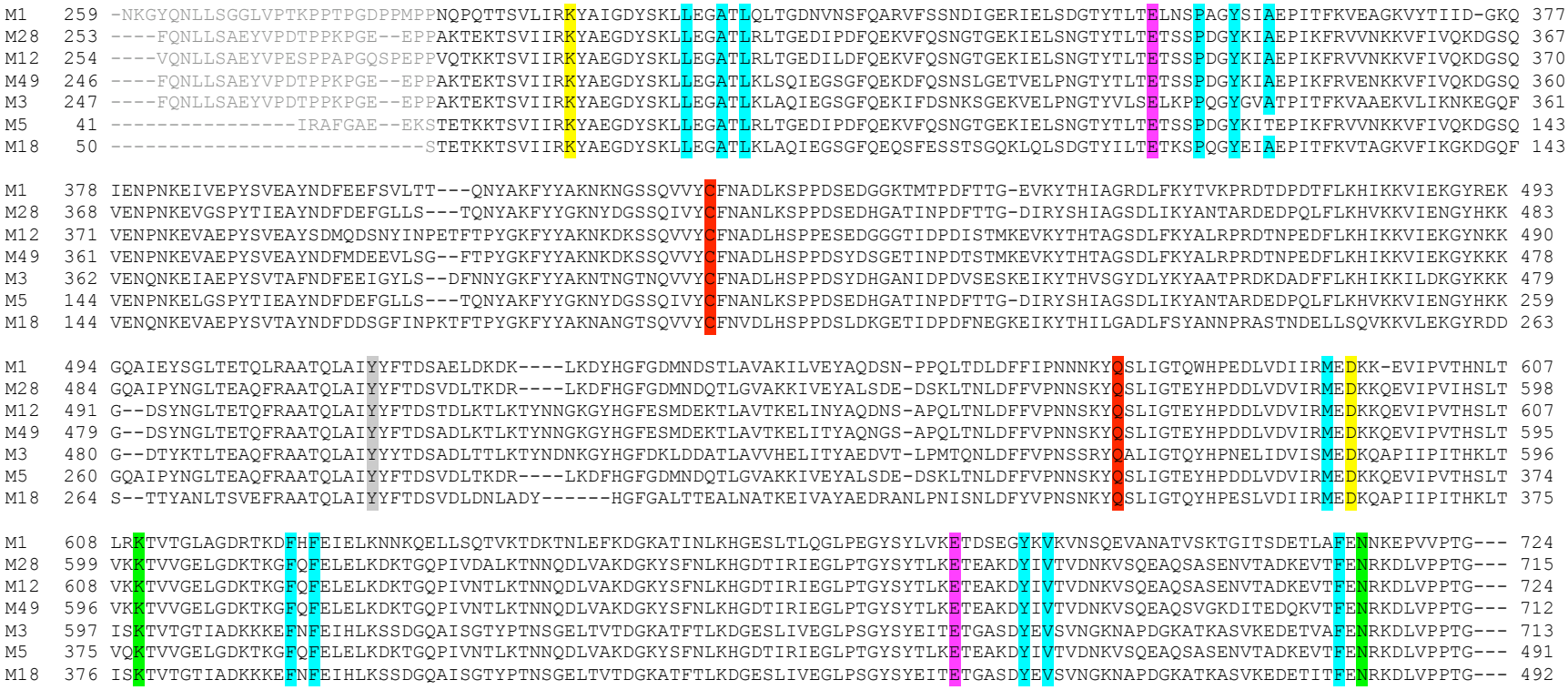
Supplemental Fig. 3 MS/MS spectra of the peptides encompassing the isopeptide bonds in Spy0125-CTR. **(A)** Peptide with m/z 774.37^{2+} containing the Lys297-Asp595 isopeptide bond. Fragment masses containing the isopeptide bond (as labeled) all include the loss of a water molecule. All fragments are in the 1+ charge state. **(B)** Peptide with m/z 743.18^{4+} containing the Lys610-Asn715 isopeptide bond. Fragment masses containing the isopeptide bond (as labeled) all include the loss of an NH_3 unit. All fragments are in the 2+ charge state. The peak of mass 1377.38 Da extends to 100 % relative abundance (the plot has been cut and zoomed to observe less abundant peaks).

Supplemental Fig. 4 Quantitation of mass spectrometry data using peptide intensity data. **(A)** The peptide containing position 575 returns a Glu ~ 7 times more frequently than Gln. **(B)** Alkylated

peptide that contains Cys426 is ~ 30 times more abundant in samples where the protein was pre-treated with methylamine (MA) and iodoacetamide (IAA) compared to iodoacetamide alone.

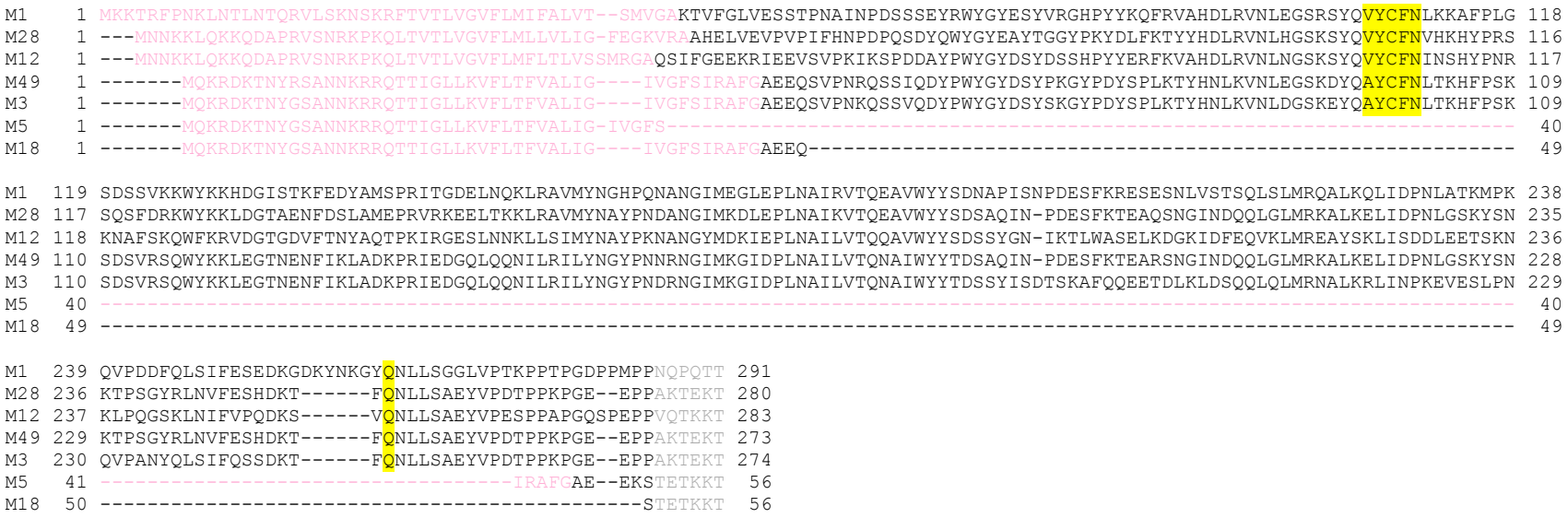
Supplemental Fig. 5 Intrinsic protein fluorescence. (**A**) Excitation and emission spectra for wild-type and Cys426Ala variant Spy0125-CTR. (**B**) Excitation and emission spectra of native (folded) and urea-unfolded Spy0125-CTR. (**C**) Emission spectra for the wild-type protein in increasing urea concentrations. The change in the spectra is shown in the inset. (**D**) Urea-dependent changes in intrinsic protein fluorescence at 360 nm for wild-type and the Cys426Ala variant.

A

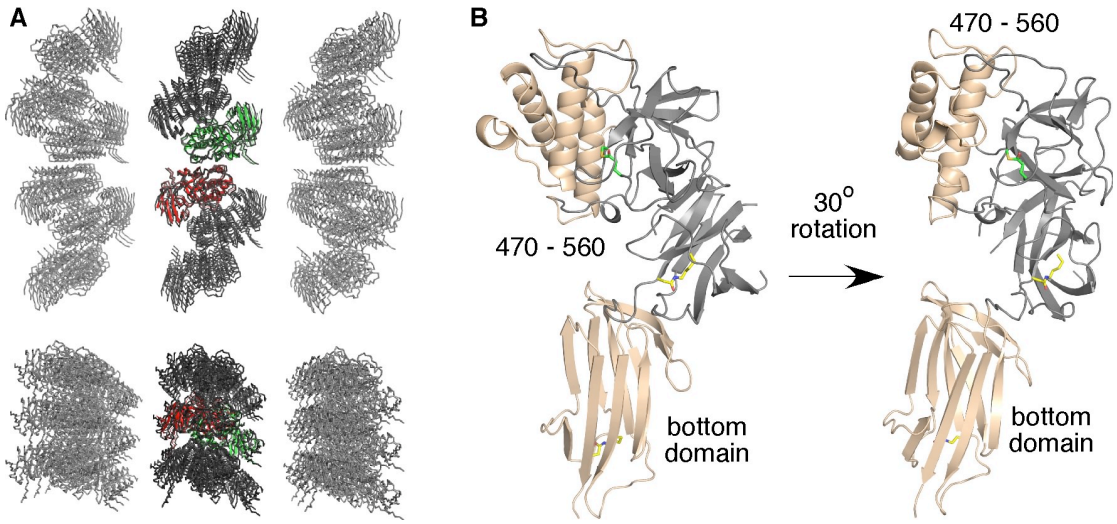


Supplemental Fig. 1

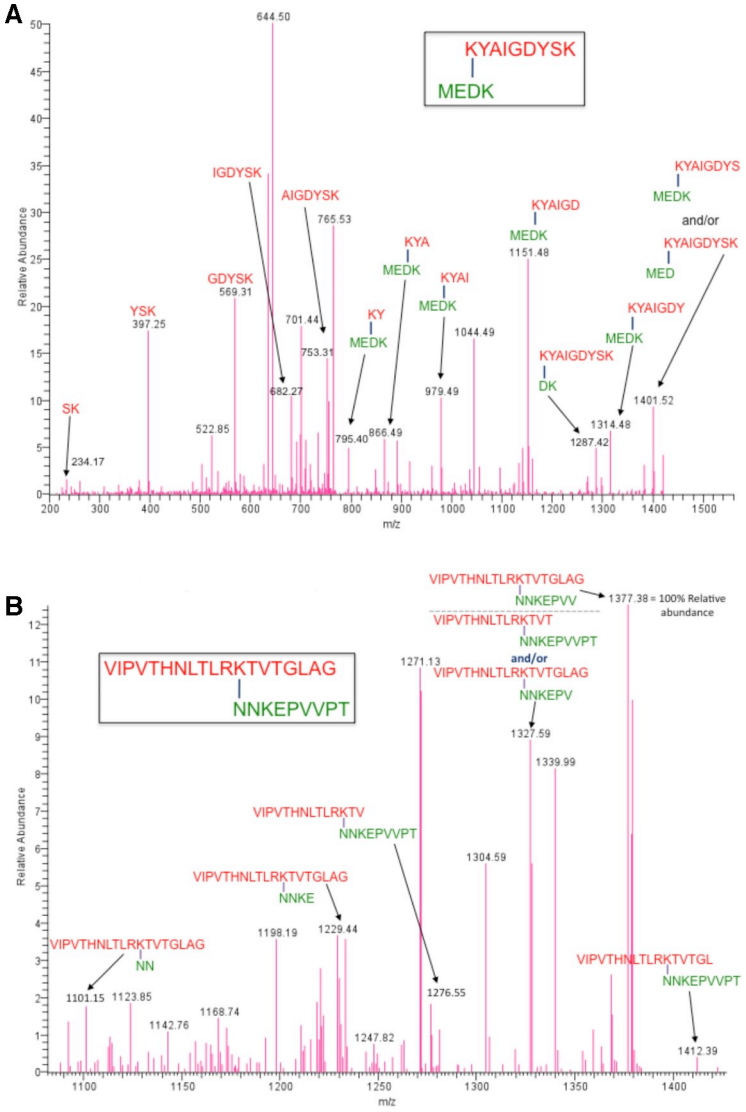
B



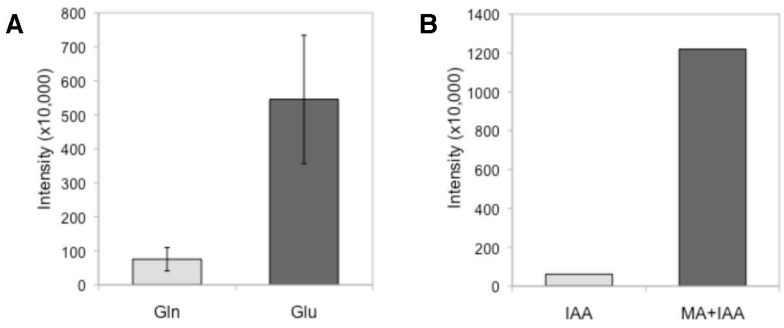
Supplemental Fig. 1



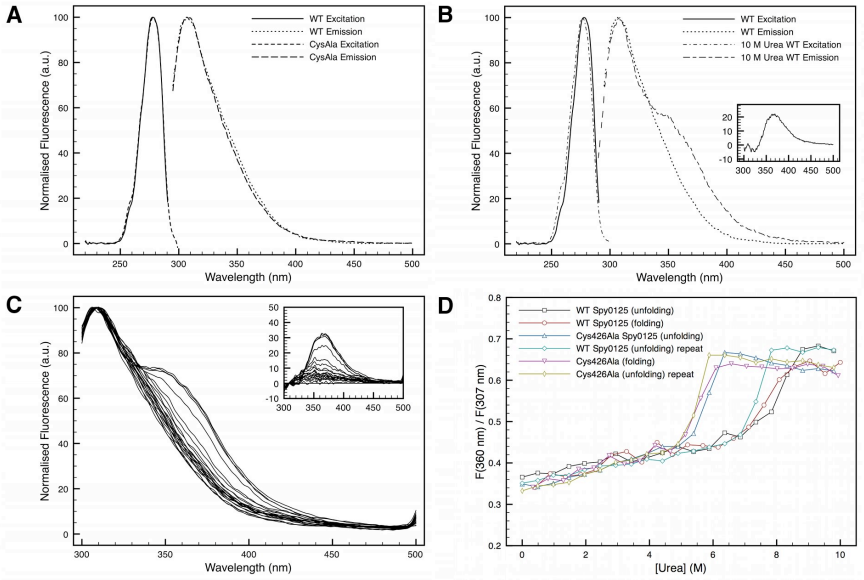
Supplemental Fig. 2



Supplemental Fig. 3



Supplemental Fig. 4



Supplemental Fig. 5

Spectrum used [a]	*Observed m/z* [b]	*Calculated m/z* [b]	$\Delta_{obs:calc}$	*Charge*	*Proposed Structure*	*Ion type*
2^+	234.17	234.15	0.02	1^+	SK	y2
2^+ and 3^+	261.18	261.09	0.09	1^+	ME	b2'
3^+	341.83	341.74	0.09	2^+	IGDYSK	y6
2^+	397.25	397.21	0.04	1^+	YSK	y3
3^+	398.28	398.26	0.02	2^+	MEDK + KY ($-H_2O$)	b6 parent
3^+	433.92	433.71	0.21	2^+	MEDK + KYA ($-H_2O$)	b7 parent
3^+	490.45	490.25	0.2	2^+	MEDK + KYAI ($-H_2O$)	b8 parent
2^+	512.25	512.24	0.01	1^+	DYSK	y4
2^+ and 3^+	569.31	569.26	0.05	1^+	GDYSK	y5
2^+	632.18	632.31	0.13	1^+	MEDK + K ($-H_2O$)	b5 parent
2^+ and 3^+	682.27	682.34	0.07	1^+	IGDYSK	y6
2^+ and 3^+	753.31	753.38	0.07	1^+	AIGDYSK	y7
2^+ and 3^+	795.40	795.28	0.12	1^+	MEDK + KY ($-H_2O$)	b6 parent
2^+ and 3^+	866.49	866.41	0.08	1^+	MEDK + KYA ($-H_2O$)	b7 parent
2^+ and 3^+	916.34	916.44	0.10	1^+	YAIGDYSK	y8
2^+	979.49	979.50	0.01	1^+	MEDK + KYAI ($-H_2O$)	b8 parent
2^+	1036.49	1036.52	0.03	1^+	MEDK + KYAIG ($-H_2O$)	b9 parent
2^+	1151.48	1151.55	0.07	1^+	MEDK + KYAIGD ($-H_2O$)	b10 parent
2^+	1287.42	1287.66	0.24	1^+	KYAIGDYSK + DK ($-H_2O$)	y11 parent
2^+	1314.48	1314.61	0.13	1^+	MEDK + KYAIGDY ($-H_2O$)	b11 parent
2^+	1401.52	1401.64	0.12	1^+	MEDK + KYAIGDYS ($-H_2O$)	b12 parent
2^+	1401.52	1401.64	0.18	1+	KYAIGDYSK + MED ($-H_2O$)	b3 parent
2^+	1416.57	1416.70	0.13	1+	KYAIGDYSK + EDK ($-H_2O$)	y12 parent

[a] 2^+ = (m/z) 774.37^{2+}; 3^+ = (m/z) 516.59^{3+}

[b] Monoisotopic masses.

SUPPLEMENTAL TABLE 1. *MS/MS of peptides with m/z 774.37^{2+} and 516.59^{3+} containing the Lys297-Asp595 isopeptide bond of Spy0125-CTR.*

Spectrum used [a]	*Observed m/z* [b]	*Calculated m/z* [b]	$\Delta_{obs:calc}$	*Charge*	*Proposed Structure*	*Ion type*
4^+	213.08	213.16	0.08	1^+	VI	b2
4^+	217.13	217.12	0.01	1^+	PT	y2’
3^+ and 4^+	316.14	316.19	0.05	1^+	VPT	y3’
3^+ and 4^+	418.11	418.23	0.12	1^+	TGLAG	y5
4^+	510.22	510.33	0.11	1^+	VIPVT	b5
3^+ and 4^+	512.40	512.31	0.09	1^+	PVVPT	y5’
4^+	647.44	647.39	0.05	1^+	VIPVTH	b6
3^+ and 4^+	761.43	761.43	0.00	1^+	VIPVTHN	b7
3^+ and 4^+	874.37	874.52	0.15	1^+	VIPVTHNL	b8
3^+	1101.15	1101.16	0.01	2^+	VIPVTHNLTLRKTVTGLAG + NN ($-NH_3$)	b21’ parent
3^+	1229.44	1229.73	0.29	2^+	VIPVTHNLTLRKTVTGLAG + NNKE ($-NH_3$)	b23’ parent
3^+	1276.55	1276.77	0.22	2^+	VIPVTHNLTLRKTV + NNKEPVVPT ($-NH_3$)	b14 parent
3^+	1327.59	1327.29	0.30	2^+	VIPVTHNLTLRKTVT + NNKEPVVPT ($-NH_3$)	b15 parent
3^+	1327.59	1327.79	0.20	2^+	VIPVTHNLTLRKTVTGLAG + NNKEPV ($-NH_3$)	b25’ parent
3^+	1377.38	1377.32	0.06	2^+	VIPVTHNLTLRKTVTGLAG + NNKEPVV ($-NH_3$)	b26’ parent
3^+	1412.39	1412.34	0.04	2^+	VIPVTHNLTLRKTVTGLAGL + NNKEPVVPT ($-NH_3$)	b17 parent

[a] 3^+ = (m/z) 990.57^{3+}; 4^+ = (m/z) 743.18^{4+}

[b] Monoisotopic masses.

SUPPLEMENTAL TABLE 2: *MS/MS of peptides with m/z* 990.57^{3+} *and* 743.18^{4+} *containing the Lys610-Asn715 isopeptide bond of Spy0125-CTR.*

Spy0125 Sample	*Proposed peptide Sequence*	*Observed m/z* [a]	*Mass (Da, experimental)* [a]	*Mass (Da, calculated)* [a]	$\Delta_{obs:calc}$	*Mascot score*	*P value from Mascot*
Wild-type	Y*E*SLIGTQWHPEDLVDIIR	1142.59^{2+}	2283.15	2283.15	0.00	81.0	8.0×10^{-9}
Incubation with methylamine	Y*Q*SLIGTQWHPEDLVDIIR	1149.10^{2+}	2296.18	2296.19	0.01	95.1	3.1×10^{-10}
Incubation with methylamine and iodoacetamide	NGSSQVVY*C*FNADLK	851.40^{2+}	1700.79	1700.78	0.01	89.0	1.30×10^{-9}
Incubation with DTT and iodoacetamide (prior to digest)	NGSSQVVY*C*FNADLK	851.40^{2+}	1700.79	1700.78	0.01	99.3	1.2×10^{-10}
Incubation with DTT (digested peptides)	NGSSQVVYCFNADLK	822.89^{2+}	1643.76	1643.76	0.00	103.6	4.3×10^{-11}
Incubation with DTT and iodoacetamide (digested peptides)	NGSSQVVY*C*FNADLK	851.40^{2+}	1700.78	1700.78	0.00	80.1	9.7×10^{-9}

[a] Monoisotopic masses.

SUPPLEMENTAL TABLE 3: *Peptides with various modifications at the position of the thioester in the intact protein, as detected by mass spectrometry analysis. Residues in italics and underlined are those modified from the wild-type protein sequence.*