Supporting Information

Top-Down Mass Spectrometry and Assigning Internal Fragments for Determining Disulfide Bond Positions in Proteins

Benqian Wei[1], Muhammad A. Zenaidee[1,3], Carter Lantz[1], Brad J. Williams[4], Sarah Totten[4], Rachel R. Ogorzalek Loo[1], Joseph A. Loo [1,2]

[1] Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, CA USA

[2] Department of Biological Chemistry, University of California Los Angeles, Los Angeles, CA USA

[3] Australian Proteome Analysis Facility, Macquarie University, Macquarie Park, NSW, Australia

[4] Waters Corporation, Milford, MA, USA

Corresponding Author

Joseph A. Loo

University of California-Los Angeles, Los Angeles, CA, United States

Email: jloo@chem.ucla.edu

Table of Contents

Table S1. Unlocalized modifications for fragment matching of ExD of beta-lactoglobulin. S-3

Table S2. Unlocalized modifications for fragment matching of CAD of beta-lactoglobulin. S-4

Table S3. Unlocalized modifications for fragment matching of ExD of lysozyme, ribonuclease A, and alpha-lactalbumin. S-5

Table S4. Unlocalized modification for fragment matching of CAD of lysozyme, ribonuclease A, and alpha-lactalbumin. S-6

Table S5. Localized modifications for fragment matching of CAD of trypsin inhibitor to determine the disulfide connectrivity. S-7

Table S6. Localized modifications for fragment matching of CAD of alpha lactalbumin to determine the disulfide connectivity. S-7

Table S7. Localized modifications for fragment matching of CAD of lysozyme to determine the disulfide connectivity. S-7

Fig. S1. Representative ECD and CAD MS/MS spectra of betalactoglobulin, $[Blac + 14+]^{14+}$ and their corresponding fragmentation location maps. S-8

Fig. S2. Representative ECD and CAD MS/MS spectra of lysozyme, $[Lys + 10+]^{10+}$ and their corresponding fragmentation location maps. S-9

Fig. S3. The extent of sequence information obtained by terminal and internal fragments for ribonuclease A at different sequence regions. S-10

Fig. S4. The extent of sequence information obtained by terminal and internal fragments for alpha-lactalbumin at different sequence regions. S-11

Fig. S5. Number of disulfide bonds cleaved by terminal and internal fragments for ribonuclease A S-12

Fig. S6. Number of disulfide bonds cleaved by terminal and internal fragments for alpha-lactalbumin. S-13

Fig. S7. ECD MS/MS data of beta-lactoglobulin, $[Blac + 15H]^{15+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer. S-14

Fig. S8. ECD MS/MS data of lysozyme, $[Lys + 9H]^{9+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer S-15

Fig. S9. Representative internal fragments traversing intact disulfide bonds generated by CAD of trypsin inhibitor, $[TI + 17H]^{17+}$. S-16

Fig. S10. Representative internal fragments traversing the two intact disulfide bonds in the interior of alpha-lactalbumin generated by CAD of alpha-lactalbumin. S-17

Fig. S11. Fragment location map after importing the localized modification of CAD of lysozyme and representative internal fragmnts traversing the two intact disulfide bonds located in the interior of lysozyme generated by CAD. S-18

Table S1. Unlocalized modifications for fragment matching of ExD of β-lactoglobulin (2 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
H+	1.00783	Any	1
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	4
4H-	-4.03132	C	5
1S	31.9721	C	1
2S	63.9442	C	2
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
S-	-31.9721	C	1
2S-	-63.9442	C	2
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
SH2-	-33.9878	C	1
2(SH2-)	-67.9755	C	2

Table S2. Unlocalized modifications for fragment matching of CAD of β-lactoglobulin (2 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
2H2O-	-36.02113	Any	0
H2O-	-18.010565	Any	0
2NH3-	-34.053098	Any	0
NH3-	-17.026549	Any	0
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	4
4H-	-4.03132	C	5
1S	31.9721	C	1
2S	63.9442	C	2
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
S-	-31.9721	C	1
2S-	-63.9442	C	2
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
SH2-	-33.9878	C	1
2(SH2-)	-67.9755	C	2

Table S3. Unlocalized modifications for fragment matching of ExD of lysozyme, ribonuclease A, and α-lactalbumin (all with 4 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
H+	1.00783	Any	1
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	3
4H-	-4.03132	C	4
5H-	-5.03915	C	4
6H-	-6.04698	C	4
7H-	-7.05481	C	4
8H-	-8.06264	C	4
1S	31.9721	C	1
2S	63.9442	C	2
3S	95.9163	C	3
4S	127.8884	C	4
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
3SH-	-98.9397	C	3
4SH-	-131.92	C	4
S-	-31.9721	C	1
2S-	-63.9442	C	2
3S-	-95.9163	C	3
4S-	-127.888	C	4
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
3(H-S+)	92.8929	C	3
4(H-S+)	123.8572	C	4

Table S4. Unlocalized modification for fragment matching of CAD of lysozyme, ribonuclease A, and α-lactalbumin (all with 4 disulfide bonds) to account for all disulfide-containing fragment ions.

Modification type	Mass shift (Da)	Amino acid residue	Number of residues required
2H2O-	-36.0211	Any	0
H2O-	-18.0106	Any	0
2NH3-	-34.0531	Any	0
NH3-	-17.0265	Any	0
H-	-1.00783	C	1
2H-	-2.01566	C	2
3H-	-3.02349	C	3
4H-	-4.03132	C	4
5H-	-5.03915	C	4
6H-	-6.04698	C	4
7H-	-7.05481	C	4
8H-	-8.06264	C	4
1S	31.9721	C	1
2S	63.9442	C	2
3S	95.9163	C	3
4S	127.8884	C	4
SH-	-32.9799	C	1
2SH-	-65.9598	C	2
3SH-	-98.9397	C	3
4SH-	-131.92	C	4
S-	-31.9721	C	1
2S-	-63.9442	C	2
3S-	-95.9163	C	3
4S-	-127.888	C	4
H-S+	30.9643	C	1
2(H-S+)	61.9286	C	2
3(H-S+)	92.8929	C	3
4(H-S+)	123.8572	C	4

Table S5. Localized modifications for fragment matching of CAD of trypsin inhibitor, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	39
H-	-1.00783	86
H-	-1.00783	136
H-	-1.00783	145

Table S6. Localized modifications for fragment matching of CAD of α-lactalbumin, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	6
H-	-1.00783	120
H-	-1.00783	28
H-	-1.00783	111
H-	-1.00783	61
H-	-1.00783	77
H-	-1.00783	73
H-	-1.00783	91

Table S7. Localized modifications for fragment matching of CAD of lysozyme, which applies one hydrogen loss on every cysteine to suggest the integrity of the disulfide bonds.

Modification type	Mass shift (Da)	Position of modified amino acid residue
H-	-1.00783	6
H-	-1.00783	127
H-	-1.00783	30
H-	-1.00783	115
H-	-1.00783	64
H-	-1.00783	80
H-	-1.00783	76
H-	-1.00783	94

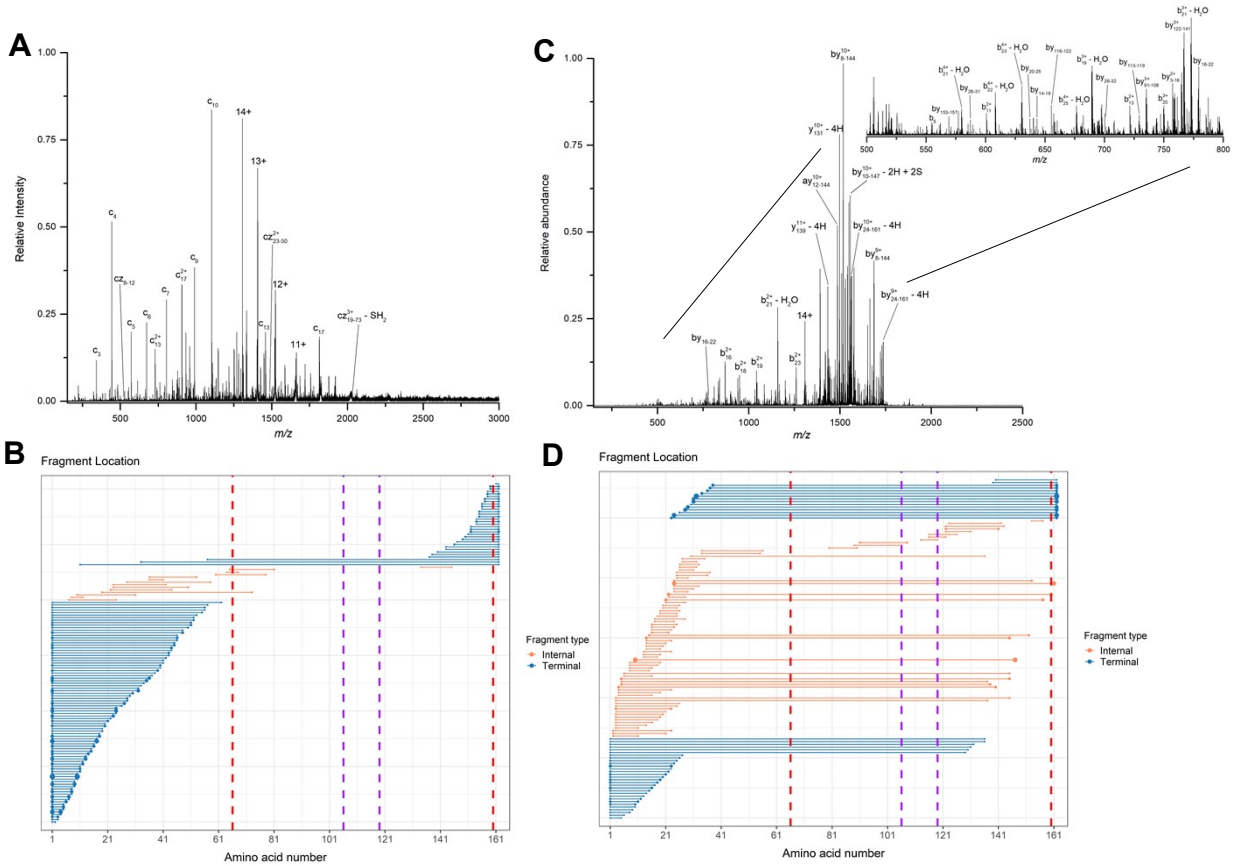

Fig. S1. A. Representative ECD MS/MS spectrum of β-lactoglobulin, $[Blac + 14H]^{14+}$. **B.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel A. **C.** Representative CAD MS/MS spectrum of β-lactoglobulin, $[Blac + 14H]^{14+}$. **D.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel C.

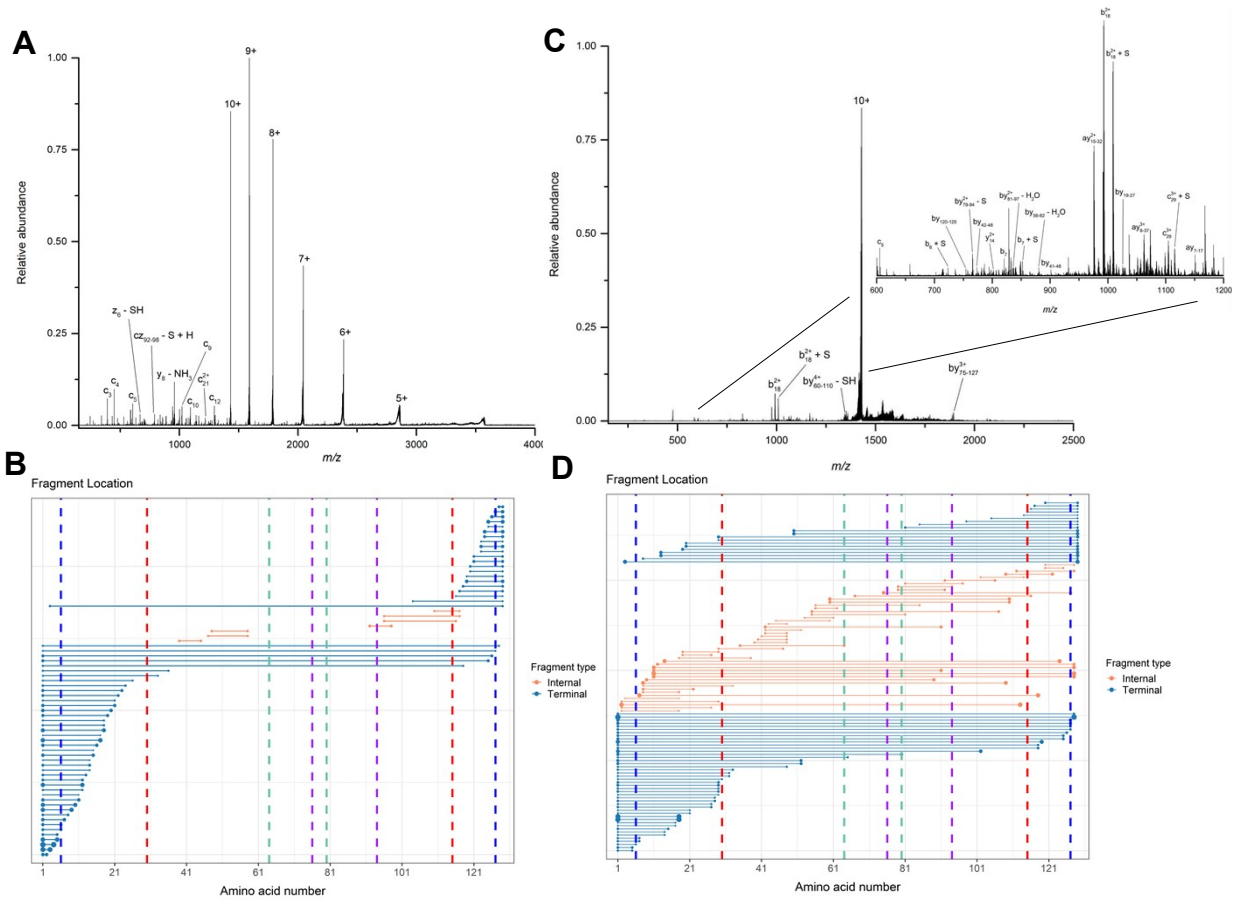

Fig. S2. A. Representative ECD MS/MS spectrum of lysozyme, $[Lys + 10H]^{10+}$. **B.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel A. **C.** Representative CAD MS/MS spectrum of lysozyme, $[Lys + 10H]^{10+}$. **D.** Fragment location map indicating the region of the protein sequence covered by terminal and internal fragments for spectrum in panel C.

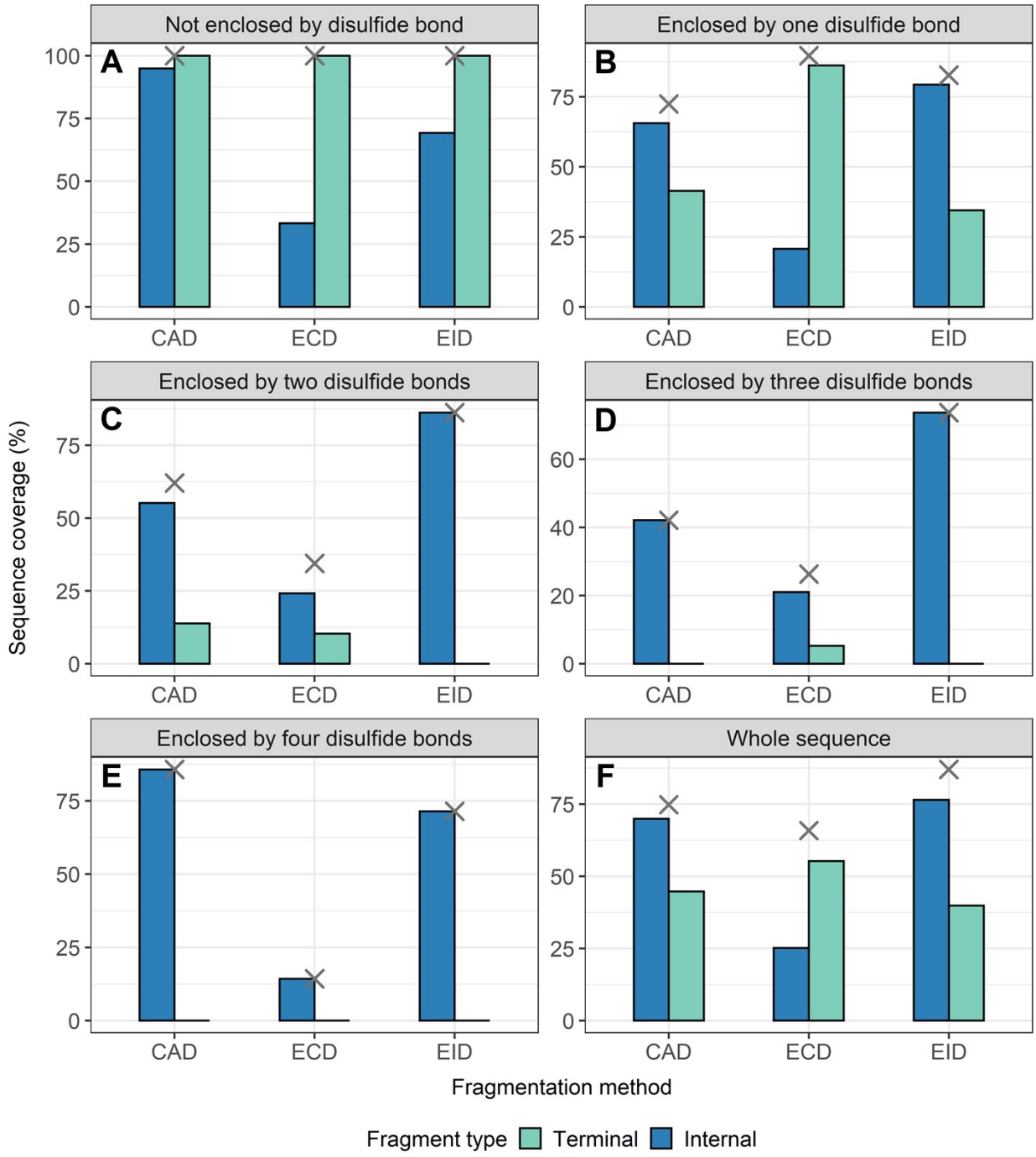

Fig. S3. The extent of sequence information obtained by terminal and internal fragments for ribonuclease A at different sequence regions after integrating data from all seven charge states (8+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** sequence not enclosed by disulfide bond, **B.** sequence enclosed by one disulfide bond, **C.** sequence enclosed by two disulfide bonds, **D.** sequence enclosed by three disulfide bonds, **E.** sequence enclosed by four disulfide bonds, **F.** whole sequence.

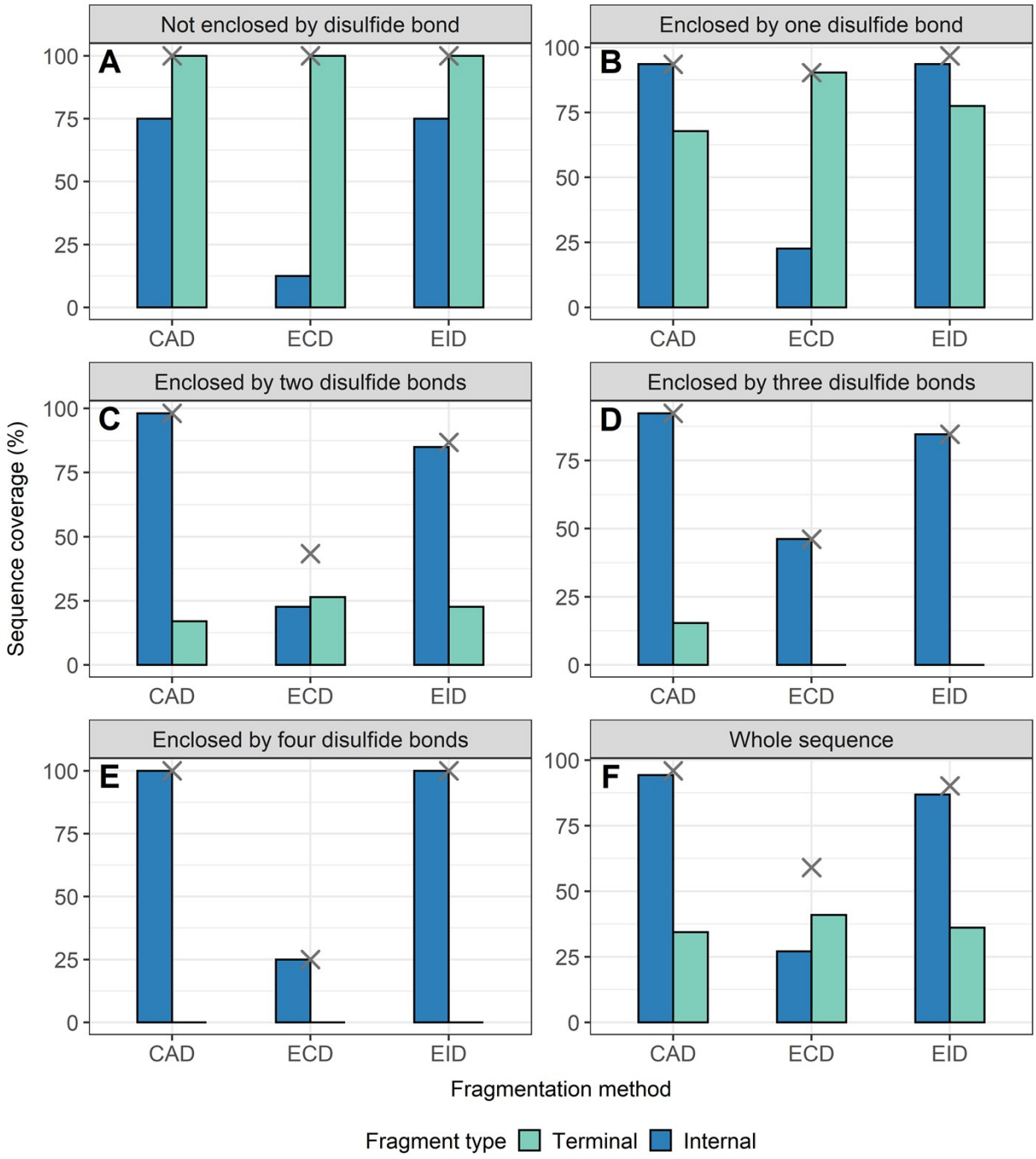

Fig. S4. The extent of sequence information obtained by terminal and internal fragments for α-lactalbumin at different sequence regions after integrating data from all four charge states (11+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** sequence not enclosed by disulfide bond, **B.** sequence enclosed by one disulfide bond, **C.** sequence enclosed by two disulfide bonds, **D.** sequence enclosed by three disulfide bonds, **E.** sequence enclosed by four disulfide bonds, **F.** whole sequence.

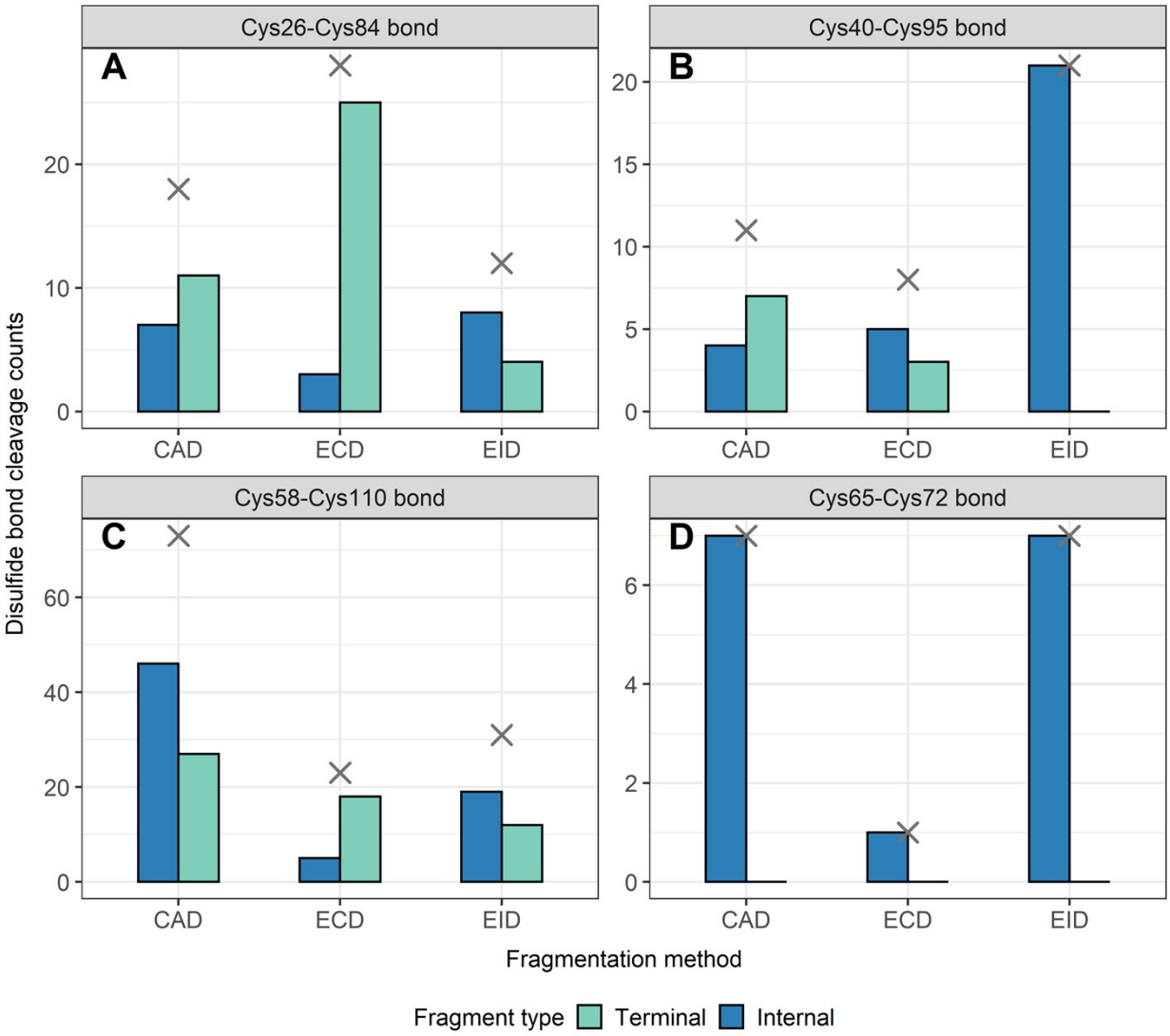

Fig. S5. Number of disulfide bonds cleaved by terminal and internal fragments for ribonuclease A after integrating data from all seven charge states (8+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** Cys26-Cys84 bond, **B.** Cys40-Cys95 bond, **C.** Cys58-Cys110 bond, **D.** Cys65-Cys72 bond.

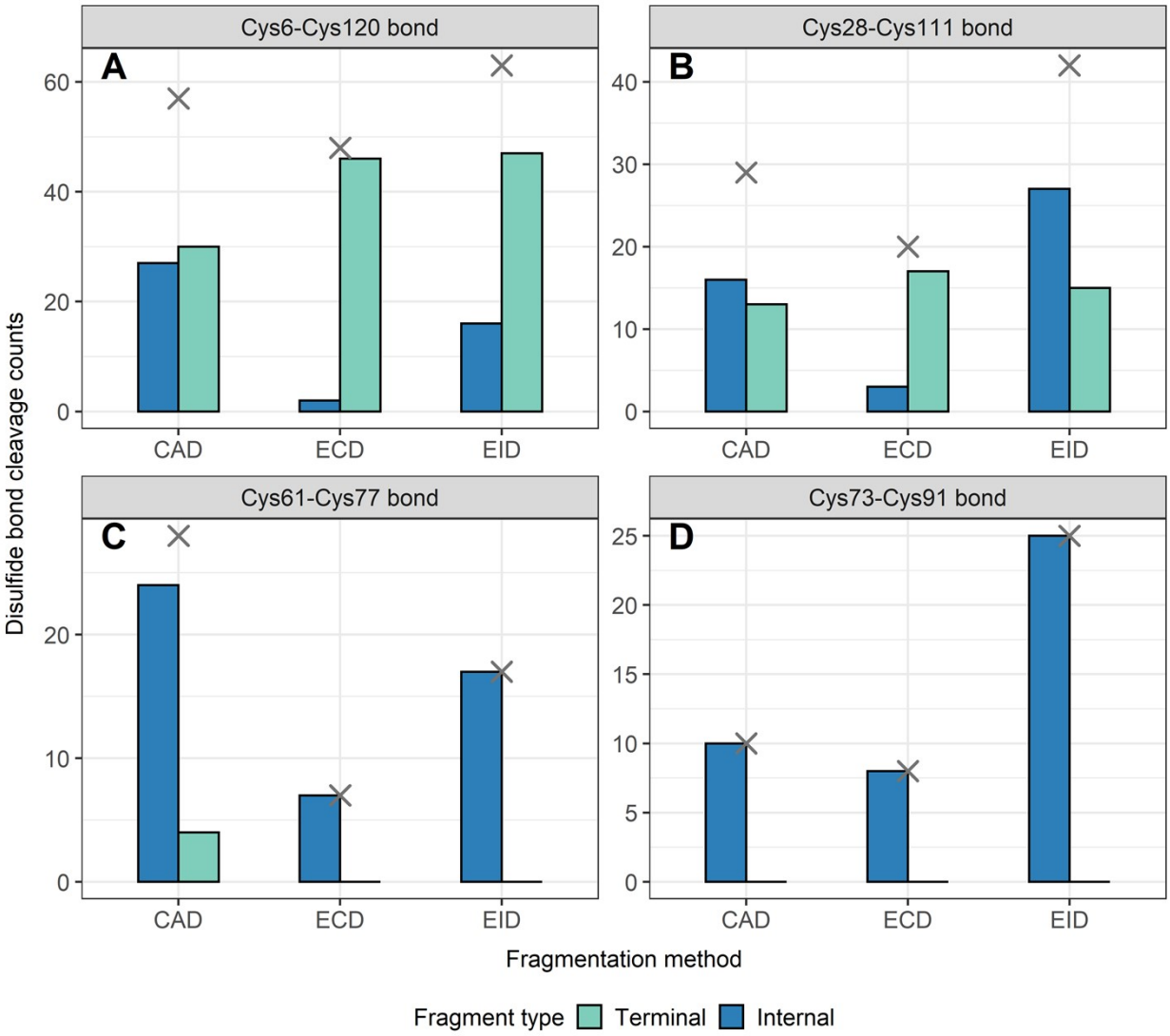

Fig. S6. Number of disulfide bonds cleaved by terminal and internal fragments for α-lactalbumin after integrating data from all four charge states (11+ to 14+) for all three fragmentation methods (CAD, ECD, and EID) examined, **A.** Cys6-Cys120 bond, **B.** Cys28-Cys111 bond, **C.** Cys61-Cys77 bond, **D.** Cys73-Cys91 bond.

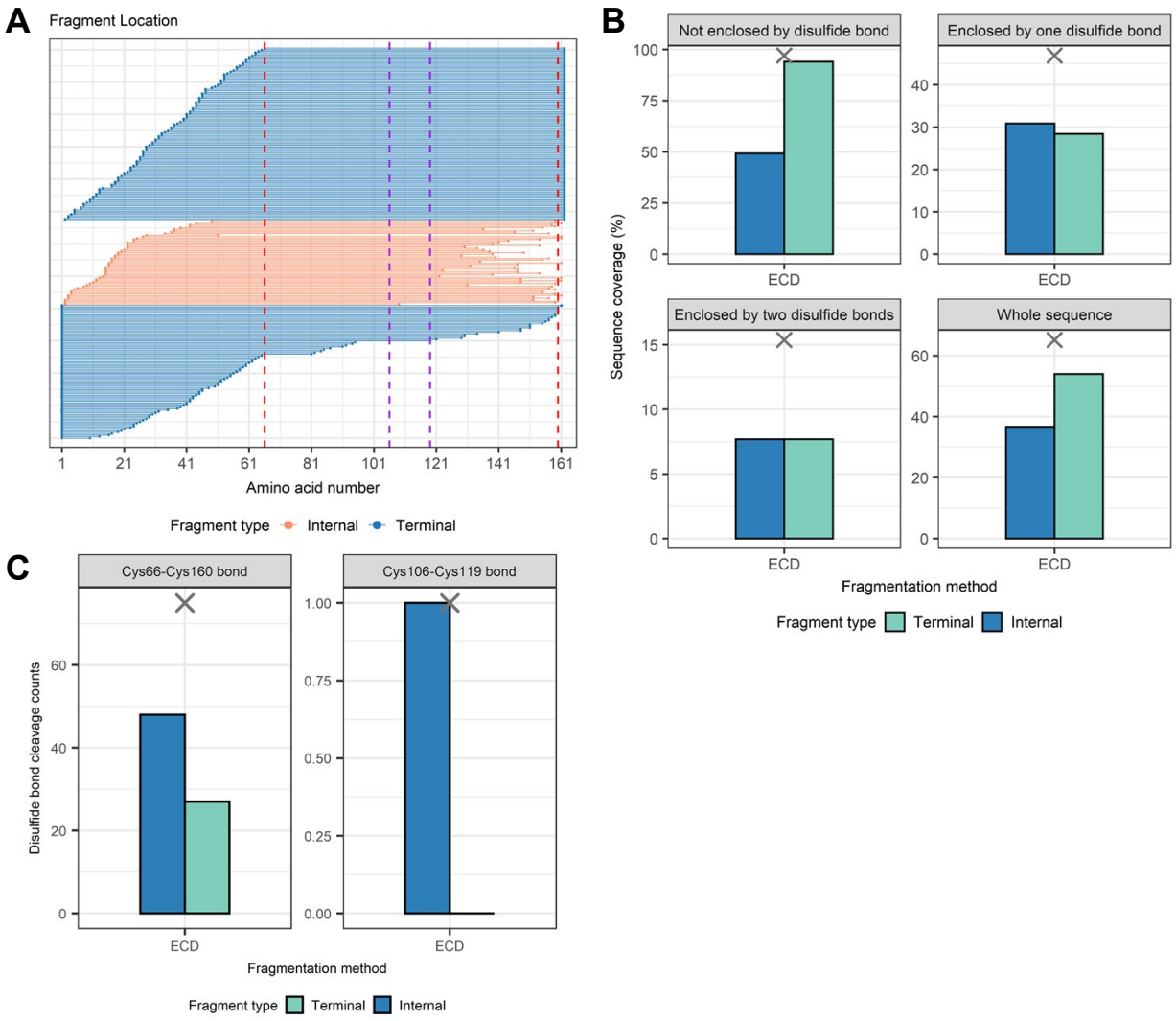

Fig. S7. ECD MS/MS data of β-lactoglobulin, $[Blac + 15H]^{15+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer, **A.** fragment location map indicating the region of the protein sequence covered by terminal and internal fragments, **B.** the extent of sequence information obtained at different sequence regions by terminal and internal fragments, **C.** the number of disulfide bonds cleaved by terminal and internal fragments. Cross markers in panels B and C indicate the metric (sequence coverage or disulfide bond cleavage counts) after combining terminal and internal fragments.

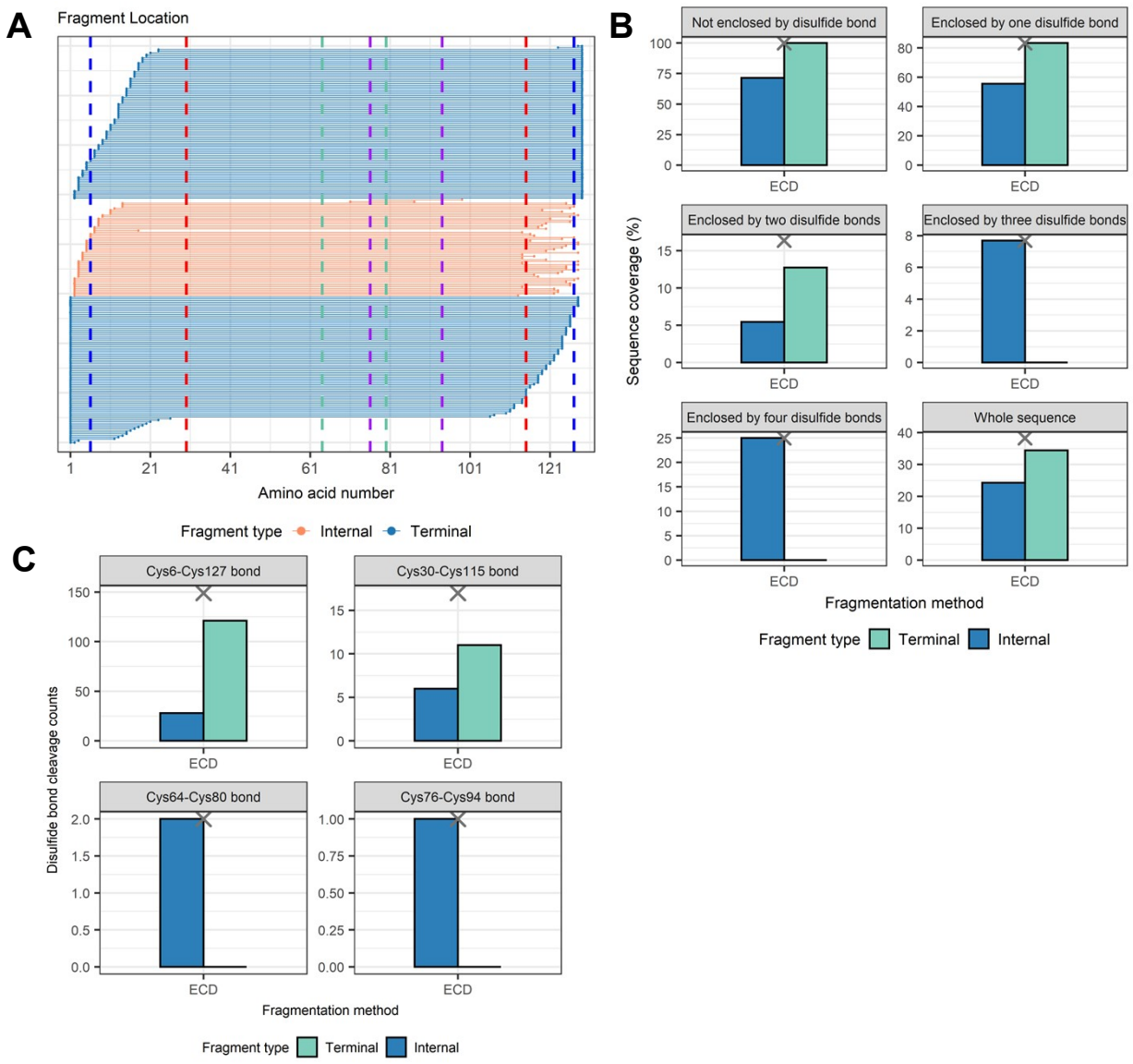

Fig. S8. ECD MS/MS data of lysozyme, $[Lys + 9H]^{9+}$ acquired from Waters Select Series Cyclic IMS mass spectrometer, **A.** fragment location map indicating the region of the protein sequence covered by terminal and internal fragments, **B.** the extent of sequence information obtained at different sequence regions by terminal and internal fragments, **C.** the number of disulfide bonds cleaved by terminal and internal fragments. Cross markers in panels B and C indicate the metric after combining terminal and internal fragments.

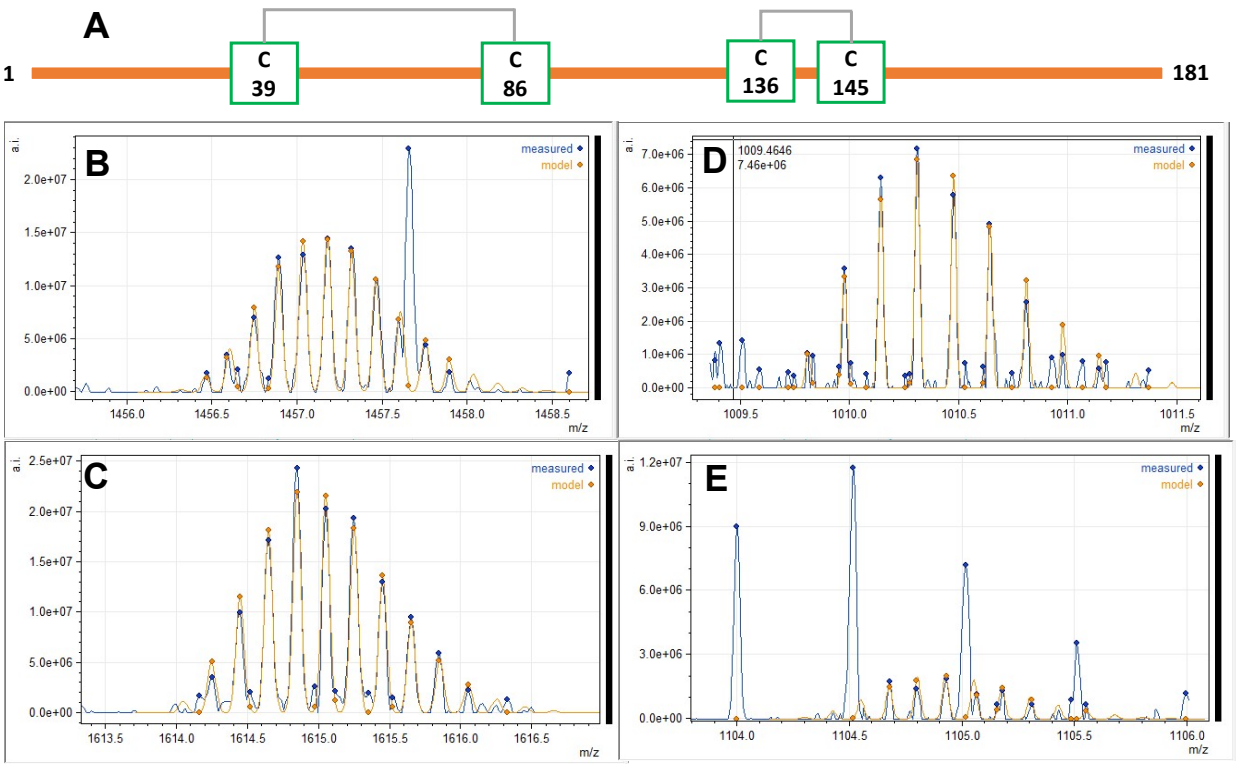

Fig. S9. A. Disulfide bond connectivities of trypsin inhibitor (2 disulfide bonds). Isotope envelope fits of representative internal fragments traversing intact disulfide bonds generated by CAD of trypsin inhibitor, $[TI + 17H]^{17+}$ (Figure 6, panel A), **B.** $[by_{4\text{-}97} + 7H]^{7+}$ (0.332 ppm), **C.** $[by_{26\text{-}99} + 5H]^{5+}$ (-0.001 ppm), **D.** $[by_{100\text{-}153} + 6H]^{6+}$ (-0.112 ppm), **E.** $[by_{98\text{-}175} + 8H]^{8+}$ (-0.189 ppm). Internal fragments shown in panels B and C traverse the Cys39-Cys86 bond (the "purple" disulfide bond), while internal fragments shown in panels D and E traverse the Cys136-Cys145 bond (the "red" disulfide bond).

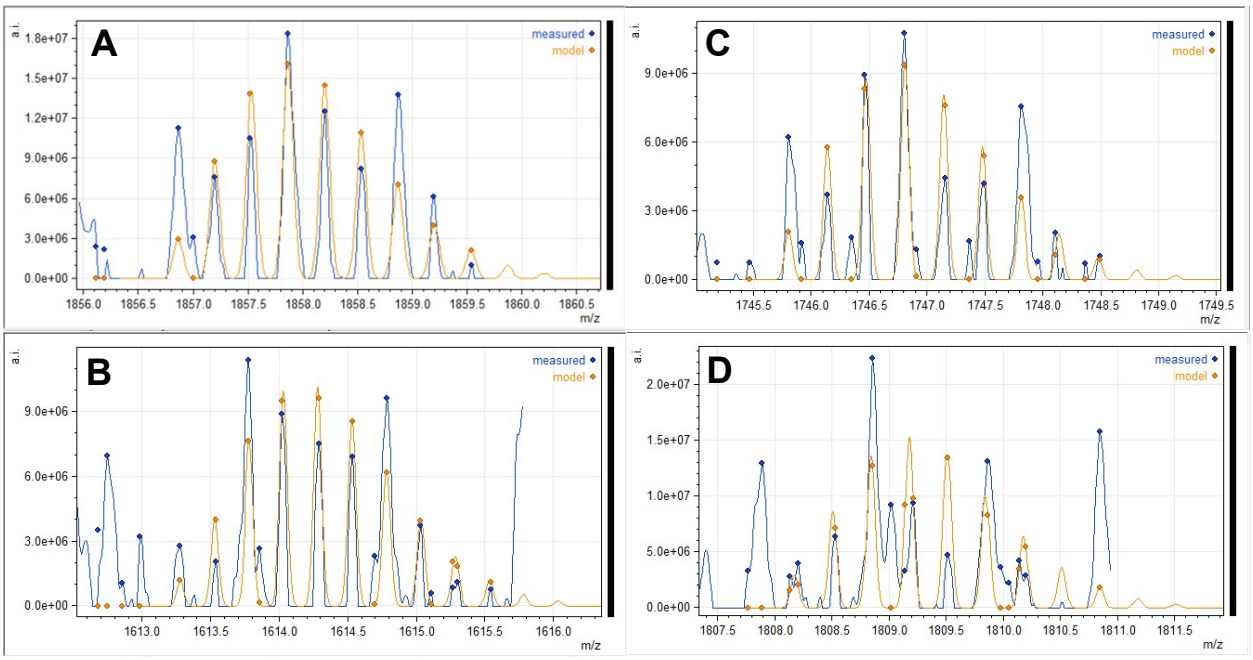

Fig. S10. Isotope envelope fits of internal fragments traversing the two intact disulfide bonds in the interior of α-lactalbumin (Cys61-Cys77 bond and Cys73-Cys91 bond, the "green" and "purple" disulfide bonds) generated by CAD (Figure 6, panel C), **A.** $[by_{50\text{-}97} + 3H]^{3+}$ (-1.205 ppm), **B.** $[by_{51\text{-}106} + 4H]^{4+}$ (-1.969 ppm), **C.** $[by_{53\text{-}97} + 3H]^{3+}$ (1.634ppm), **D.** $[by_{60\text{-}106} + 3H]^{3+}$ (1.666ppm).

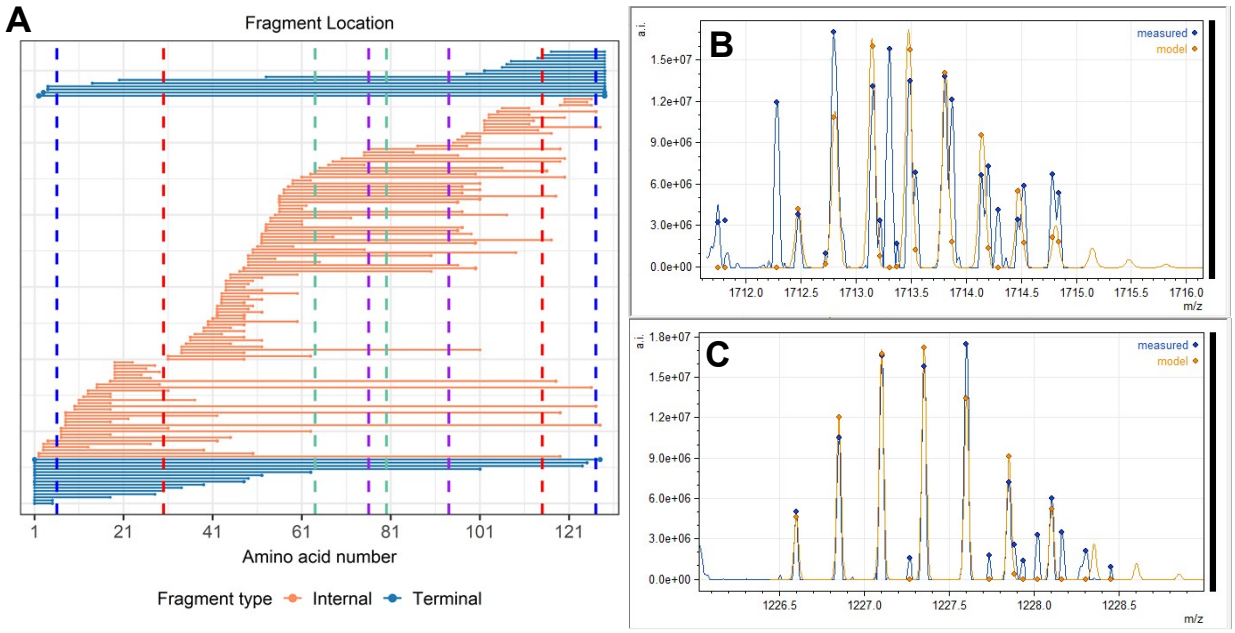

Fig. S11. A. Fragment location map after importing a hydrogen loss localized modification on every cysteine, suggesting the integrity of every disulfide bond of CAD of lysozyme after integrating data from all five charge states examined (8+ to 12+). Isotope envelope fits of representative internal fragments traversing the two intact disulfide bonds located in the interior of lysozyme (Cys64-Cys80 bond and Cys76-Cys94 bond, the "green" and "purple" disulfide bonds) generated by CAD (panel A), **B.** $[by_{49\text{-}96} + 3H]^{3+}$ (-1.579 ppm), **C.** $[by_{53\text{-}97} + 4H]^{4+}$ (-0.064 ppm).