Electronic Supplementary Material for Molecular BioSystems

Rapid Mass Spectrometric Determination of Disulfide Connectivity in Peptides and Proteins

Moitrayee Bhattacharyya[1#], Kallol Gupta[1#], Konkallu Hanumae Gowd[1$], Padmanabhan Balaram[1*]

[1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India

[#] Both the authors contributed equally

[$]Current address: Undergraduate Program, Indian Institute of Science, Bangalore-560012, India

* Address for reprint requests:

Prof. P. Balaram

Molecular Biophysics Unit, Indian Institute of Science

Bangalore – 560012, India

Fax: 91-80-236060683/ 91-23600535

Phone: 91-80-22932337

E-mail- pb@mbu.iisc.ernet.in

Supplementary Table S1: Summary of the mass spectrometric methods used for determination of disulfide connectivity

Mass spectrometric methodology for Disulfide crosslink determination	Main features	Disadvantages	DisConnect perspective
1. Morris R. *et al* (1985)	Comparison of MS profiles of the proteolytic digests of native and reduced protein identifies disulfide bonded peptides	1. Manual assignments can be prohibitively tedious for large proteins 2. Unambiguous assignments cannot be made if multiple cystines are present in a proteolytic fragment	• *DisConnect* generates specific proteolytic fragments. Non-cysteine containing fragments are matched directly against proteolytic MS profile; whereas for cysteine containing fragments, all possible combinations are queried. • Ambiguities arising from multiple structure hits for an experimental m/z value can be solved by querying its MS^2 fragment ions against the probable hits, using DisConnect-Pep. • For proteolytic fragments with multiple cystines, disulfide connectivity is determined by MS^n analysis, using *DisConnect*.
2.Yazdanparast R. *et al.* (1986, 1987)	Mass shift in gas phase Xe-assisted reduction of the proteolytic fragments of native protein identifies disulfide bonded peptides	Same as [1]	
3. Caporale C. *et al* (1996)	All possible combinations of theoretically generated linear peptide fragments, containing no more than 3 cysteines, are matched against experimental proteolytic MS profile to infer the structure of the S-S bonded peptides	1. Unambiguous assignments cannot be made if multiple cystines are present in a proteolytic fragment 2. Cannot identify a proteolytic fragment containing more than 3 cystines	
4. Fenyö D. *et al* (1997)	All possible combinations of theoretical proteolytic fragments are matched against experimental proteolytic MS profile to infer the structure of the S-S bonded peptides	1. Unambiguous assignments cannot be made if multiple cystines are present in a proteolytic fragment	
5. Craig R. et al (2003) (**Protein Disulfide Linkage Modeller**)	Implementation of a software system based on Fenyö disulfide assignment algorithm	1. Unambiguous assignments cannot be made if multiple cystines are present in a proteolytic fragment	
6. ProteinProspector (**MS-Bridge**) (**Baker, P.R. and Clauser, K.R. http://prospector.ucsf.edu**)	Theoretically generated proteolytic fragments are combined and matched against MS profile to obtain probable structures of S-S bonded peptides	1. Unambiguous assignments cannot be made if multiple cystines are present in a proteolytic fragment	
7. MS2DB (Murad et al (2011) **(MS2DB)**	MS/MS assignments of disulfide intact peptides, annotating the Biemann type ions.	1. Complete analysis of all the modes of fragmentation cannot be achieved. Hence non Biemann-type ions, arising through fragmentation at the disulfide and inside the S-S loops cannot be annotated.	
8. Gupta K. *et al* (2010)	MS^n analyses of intact disulfide bond containing native peptides provides S-S connectivity	1. Tedious manual interpretation of the data 2. Probable experimental difficulties in direct fragmentation of larger proteins	

Supplementary Figures:

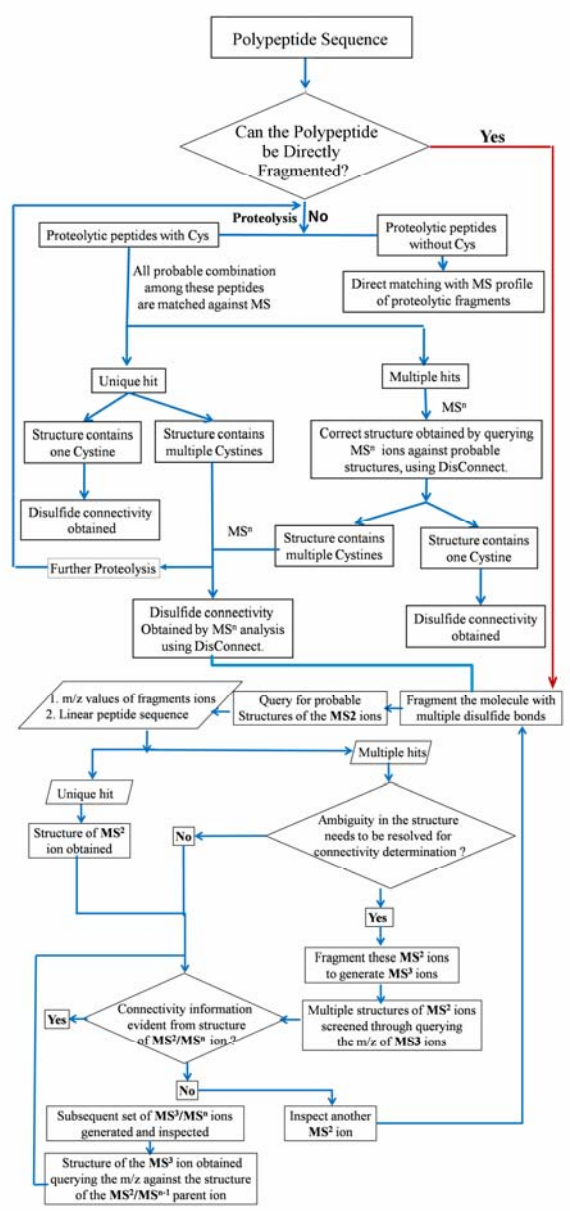

Fig. S1: Detail flowchart description of DisConnect.

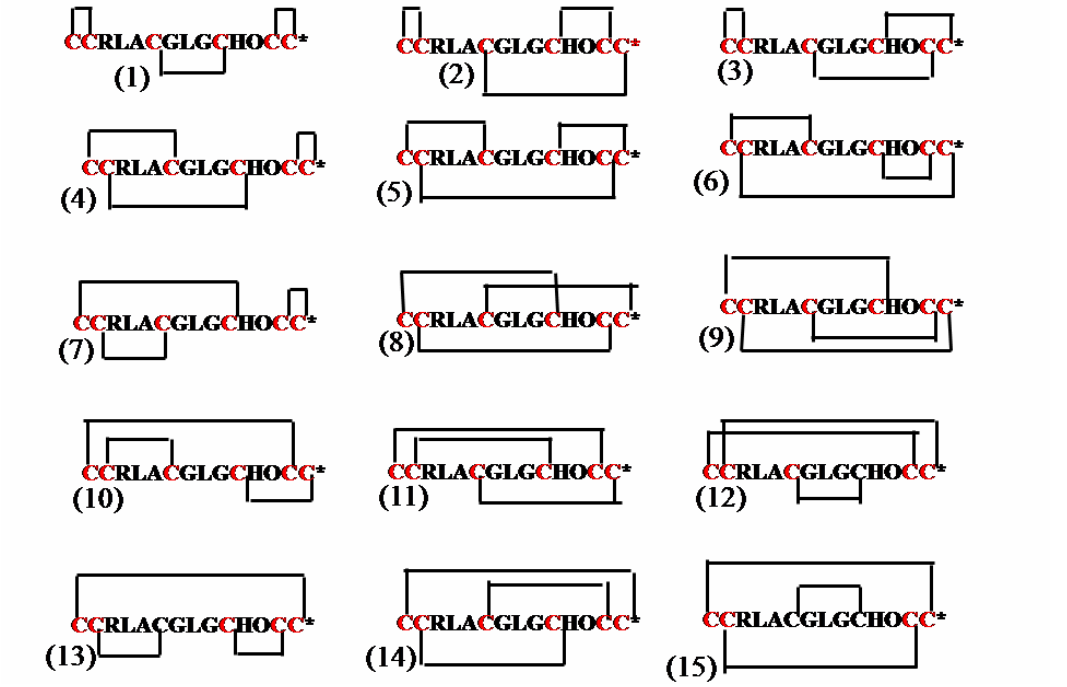

Fig. S2: 15 possible disulfide foldamers of Ar1446

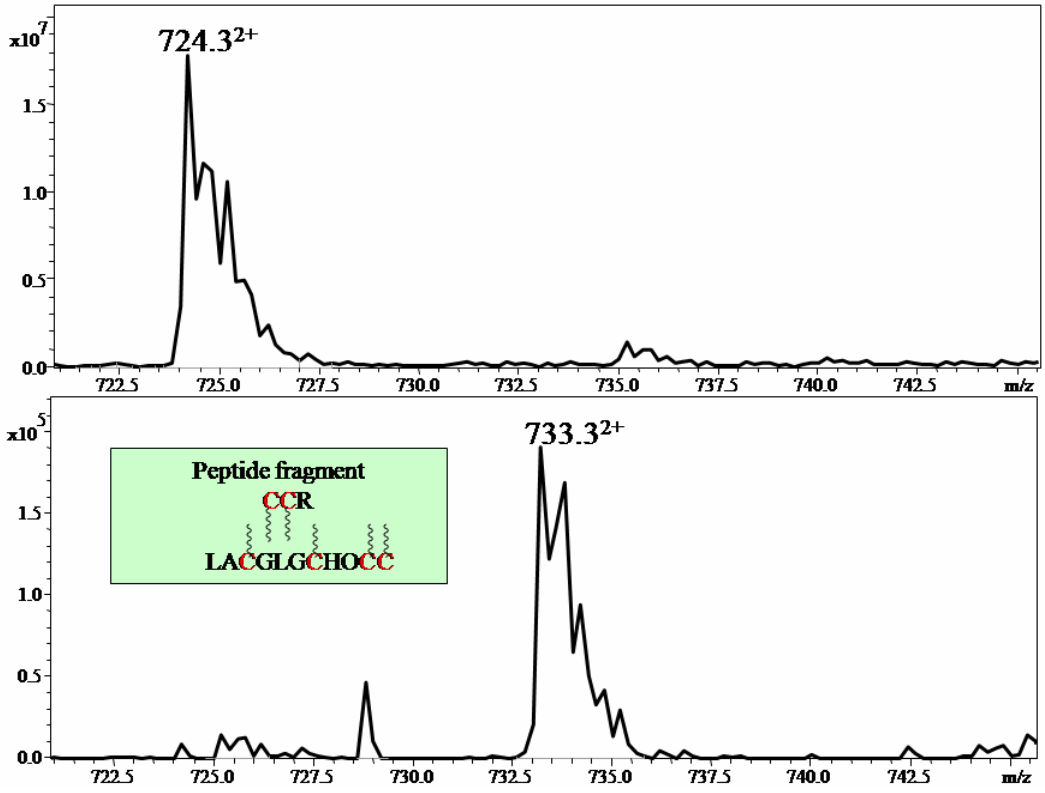

Fig. S3: Mass spectra of (a) Native and (b) trypsin digest Ar1446.

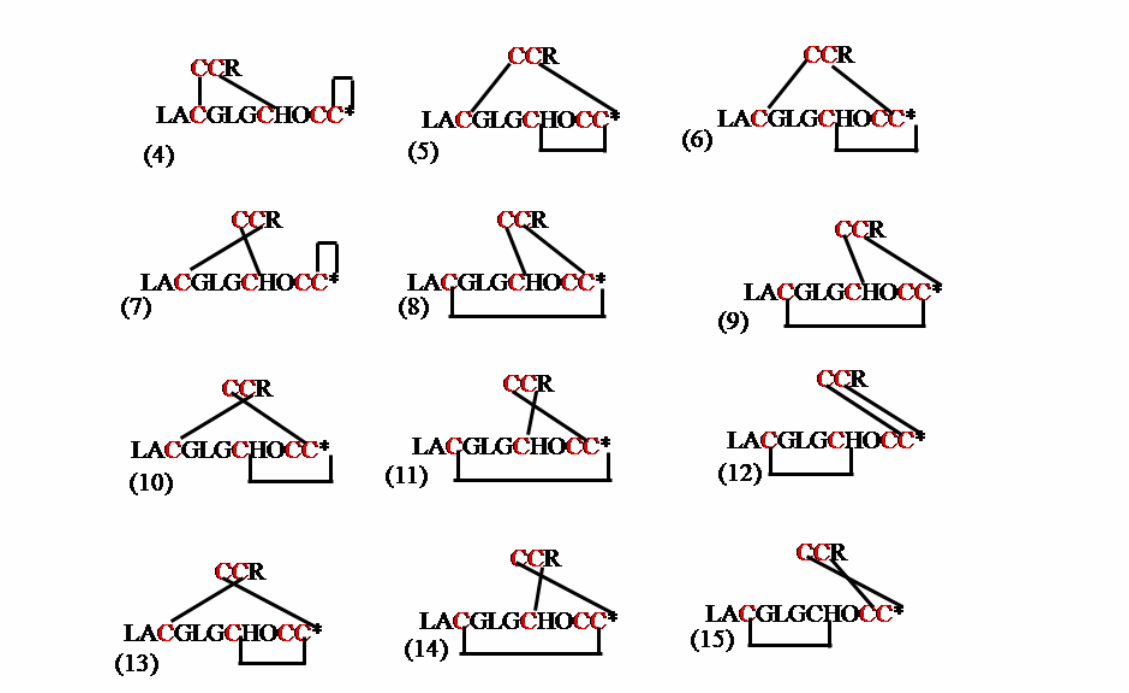

Fig. S4: 12 possible disulfide foldamers of tryptic Ar1446

622.7^{2+}
C1C2R | LAC3GLGC4HO 69|135|0|0 (I)
C1C2R | LAC3GLG | HOC5 (II)
C2R | LAC3GLGC4HOC5 69|135|0|0 (III)

676.7^{2+}
C1C2R | AC3GLGC4HOC5C6 /All Cys Connected (I)
C1C2R | LAC3GLGC4H | C5C6 /All Cys Connected (II)
C1C2R | LAC3G | GC4HOC5C6 /All Cys Connected (III)

919.2
C1C2R | LAC3 | C5C6 0|135|0|0 (I)
C1C2R | C3 | HOC5C6 69|0|0|0 (II)
C1C2R | C4HOC5C6 138|135|0|0 (III)
C1C2R | AC3 | OC5C6 0|135|0|0 (IV)

MS^3 622.7^{2+}
554.1^{2+}
C2R | LAC3GLGC4 | OC5

594.2^{2+}
C1C2R | LAC3 | LGC4HO
C1C2R | LAC3GL | C4HO

531.1
C1C2R | C6 0|135|0|0 (I)
C2R | C5C6 0|135|0|0 (II)

445.1
C1C2R 0|270|0|0

704.7^{2+}
C1C2R | LAC3GL | C4HOC5C6 /All Cys Connected (I)
C1C2R | LAC3 | LGC4HOC5C6 /All Cys Connected(II)

850.2
C1C2R | C4HO | C6 (I)
C1C2R | HOC5C6 69|135|0|0 (II)
C2R | C4HOC5C6 69|135|0|0 (III)
C2R | C3 | HOC5C6 (IV)

713.1
C1C2R | OC5C6 69|135|0|0 (I)
C2R | C3 | OC5C6 (II)
C2R | C4 | OC5C6 (III)

1328.4
C1C2R | LAC3GLGC4 | OC5C6

Fig. S5: DisConnect output for the major fragment ions present in the CID MS^2 spectrum of the tryptic Ar1446.

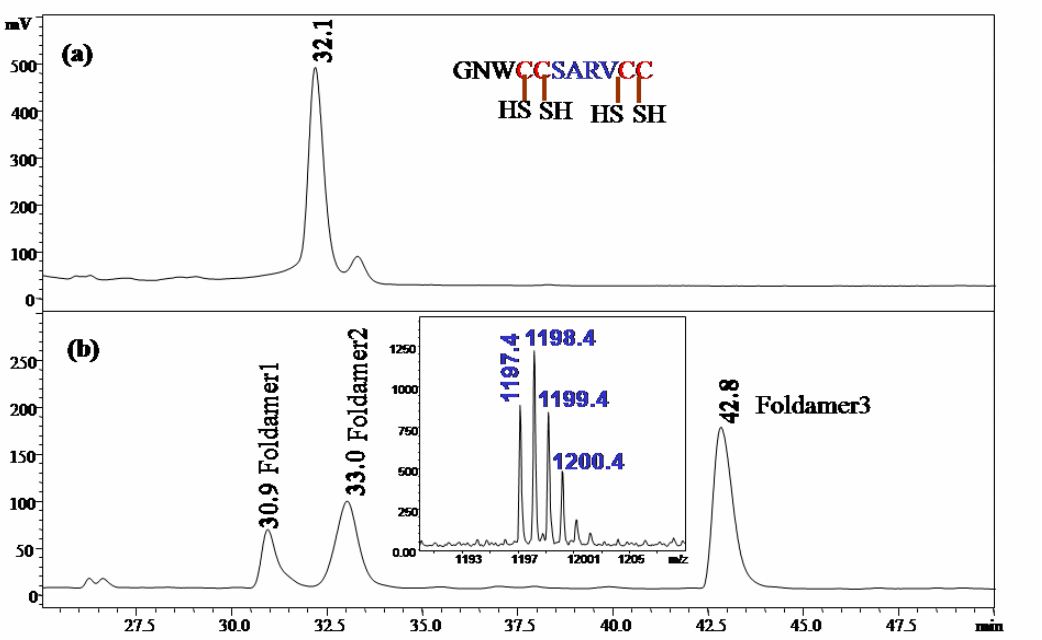

Fig. S6: (a) HPLC profile of the linear peptide. (b) HPLC profile of the oxidized mixture. Identical mass spectra for each of the HPLC fractions, corresponding to the two disulfide bonded peptide, establish the three peaks as three foldamers.

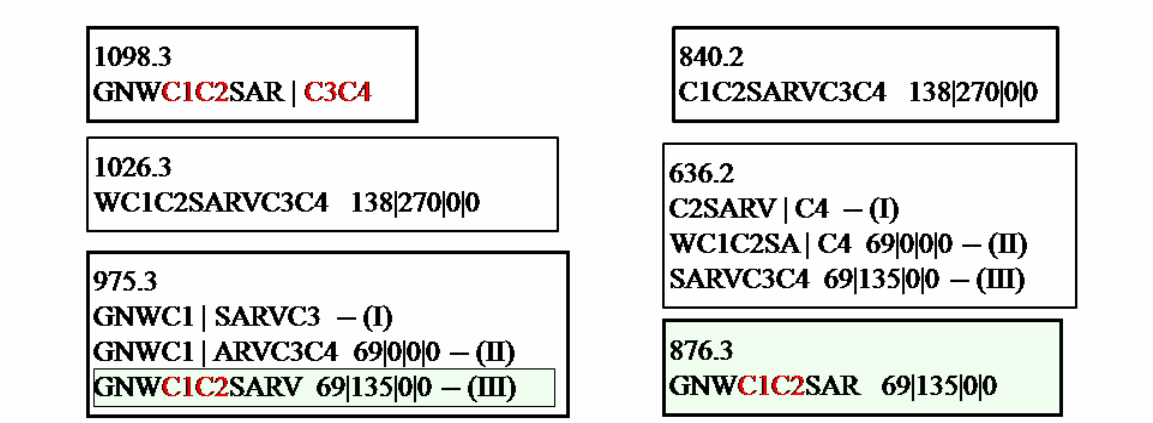

Fig. S7: DisConnect predicted structures of the key fragment ions observed in the MS^2 spectra of *des* bromo Mo1277.

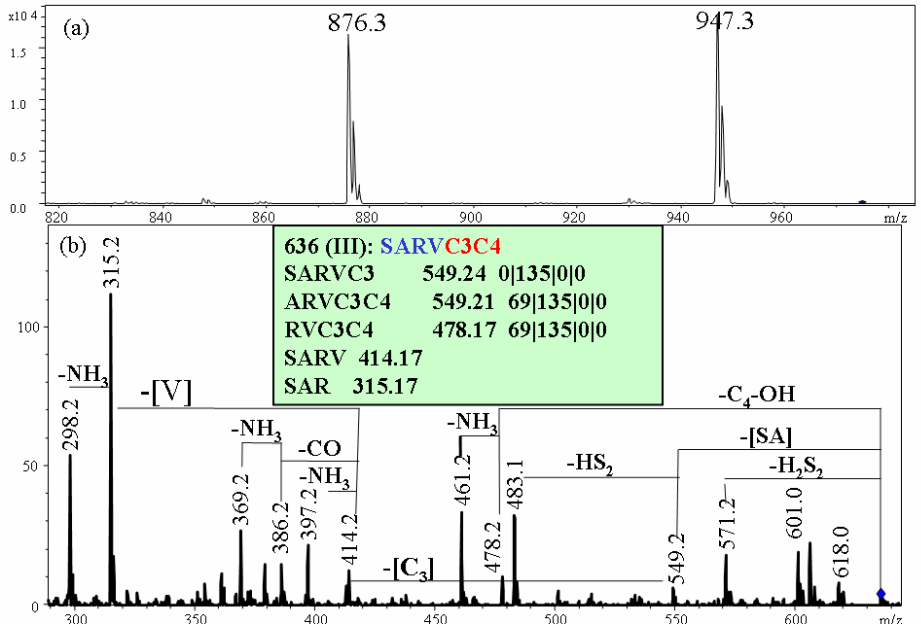

Fig. S8: (a) CID MS^3 spectrum of the ion at m/z 975.3 yielding 876.3, establishing parent product ion relationship among these ions. (b) CID MS^3 spectrum of the ion at m/z 636.1from Foldamer 1 of *des* bromo Mo1277. All the fragment ions can only be explained from structure 636.1(c).

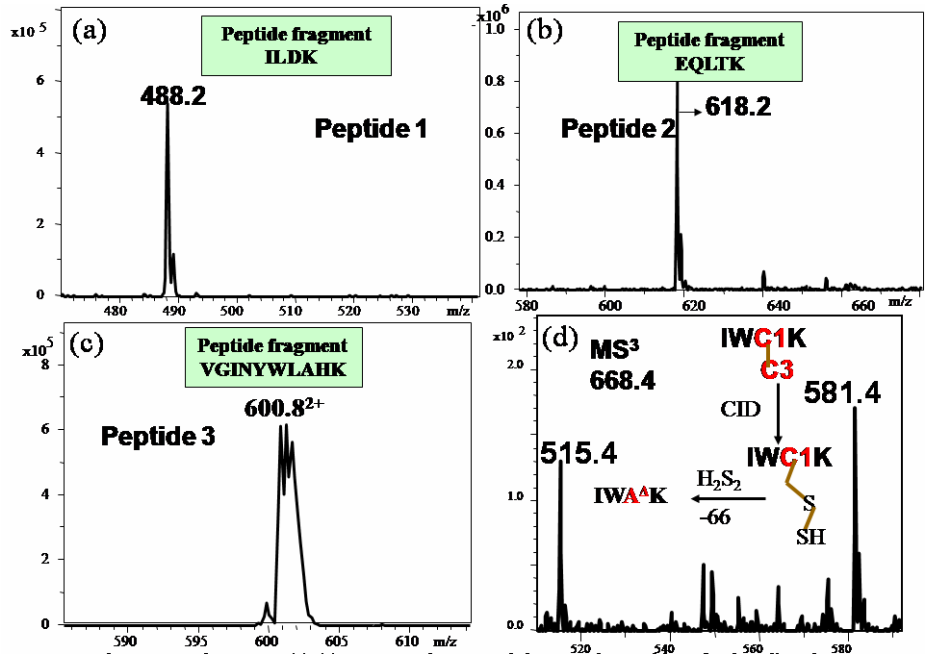

Fig. S9: (a)-(c) Non-cysteine containing tryptic peptides of α-lactalbumin. (d) MS^3 spectra of the MS^2 ion at m/z 668.4. The presence of the ions at m/z 581.4 and 515.4 confirms the structure I for 668.4.

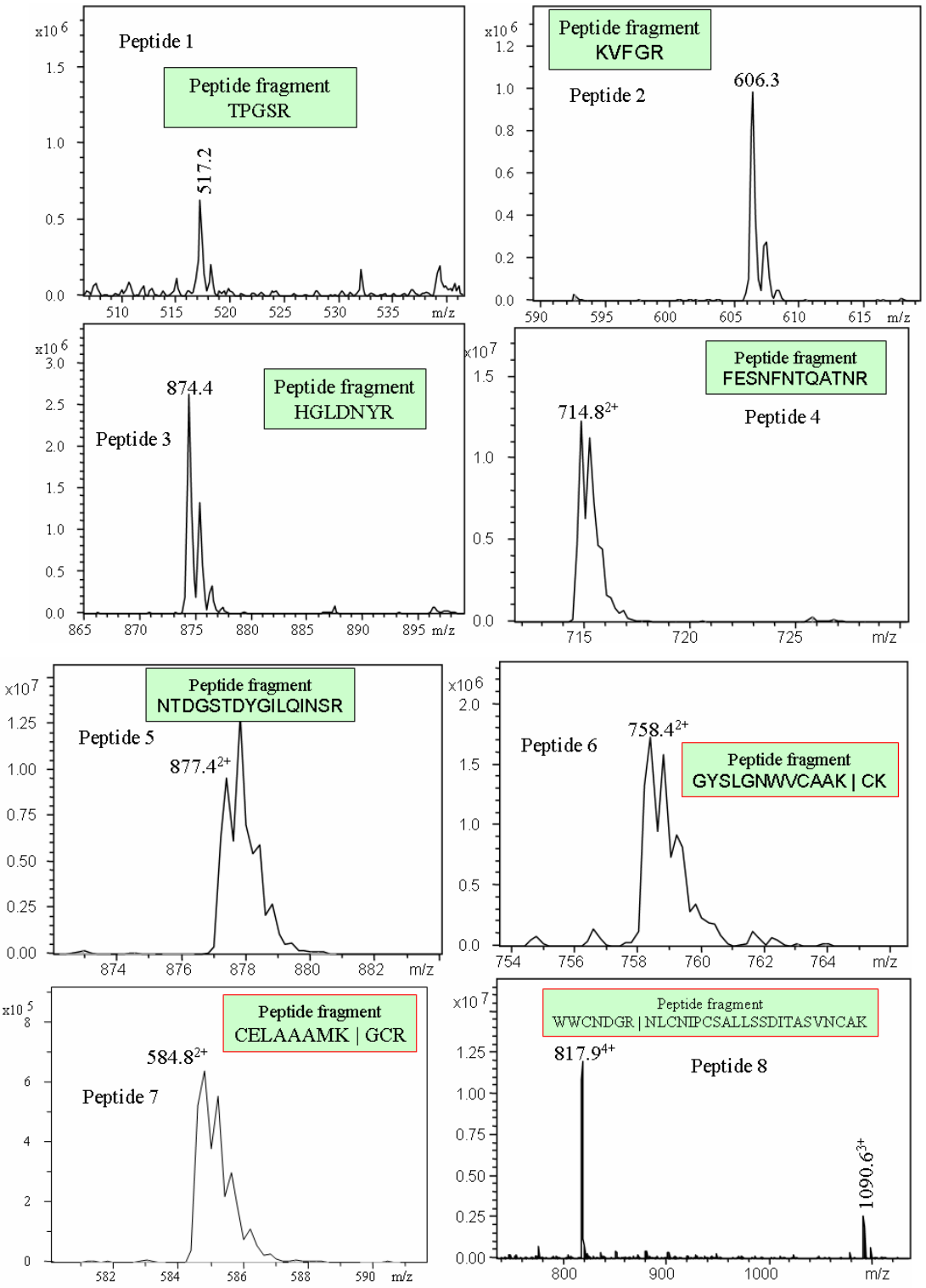

Fig. S10: Tryptic peptide 1-8 of lysozyme. Peptide 1-5 correspond to non-cysteine containing peptides. Peptide 6 and 7 contain one pairs of cysteine, whereas peptide 8 two pairs.

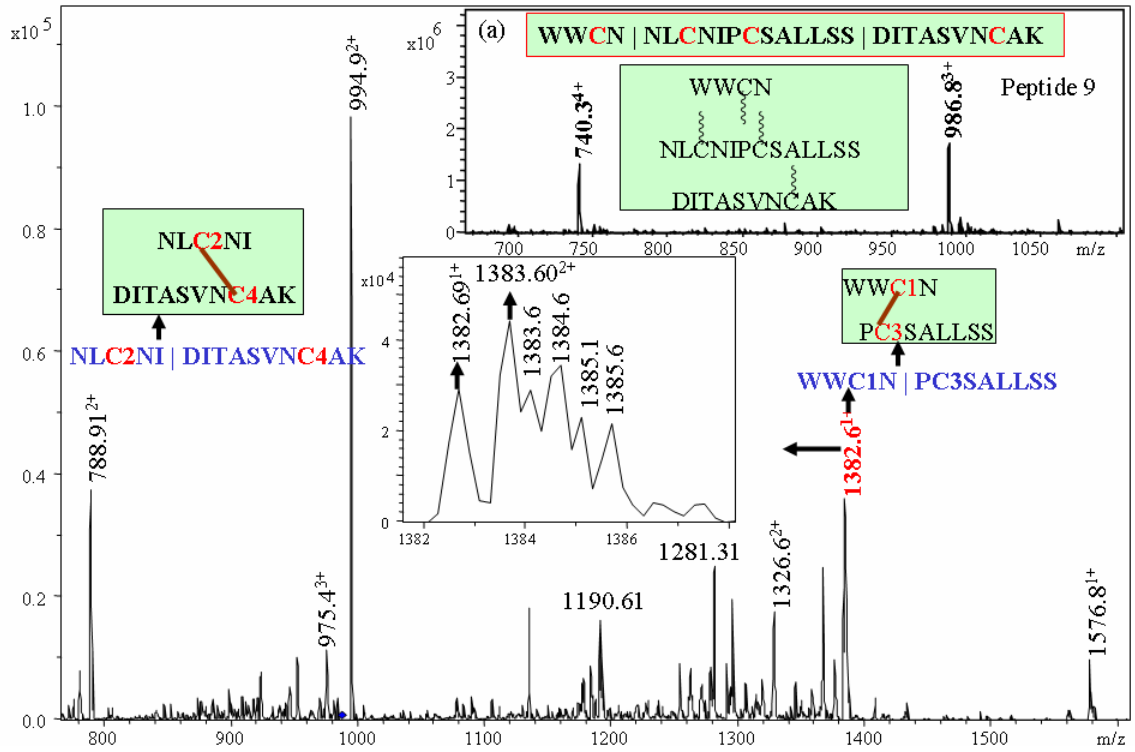

Fig. S11: CID MS^2 spectrum of peptide 9 obtained by Asp-N digestion of the tryptic digest. Inset (a) shows the MS profile of the peptide. Expansion of m/z region 1382.6 shows the co-presence of a doubly and singly charged species. The structure of the ion at m/z 1382.69 establishes C1-C3 connectivity.

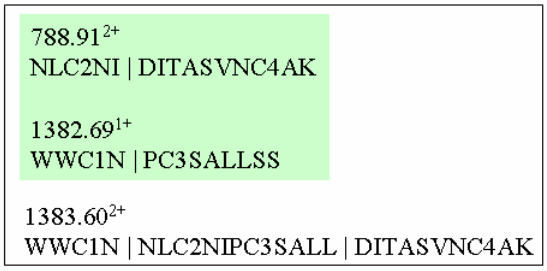

Fig. S12: DisConnect predicted possible structure/s of the major fragment ions present in the MS^2 spectrum of peptide 9.

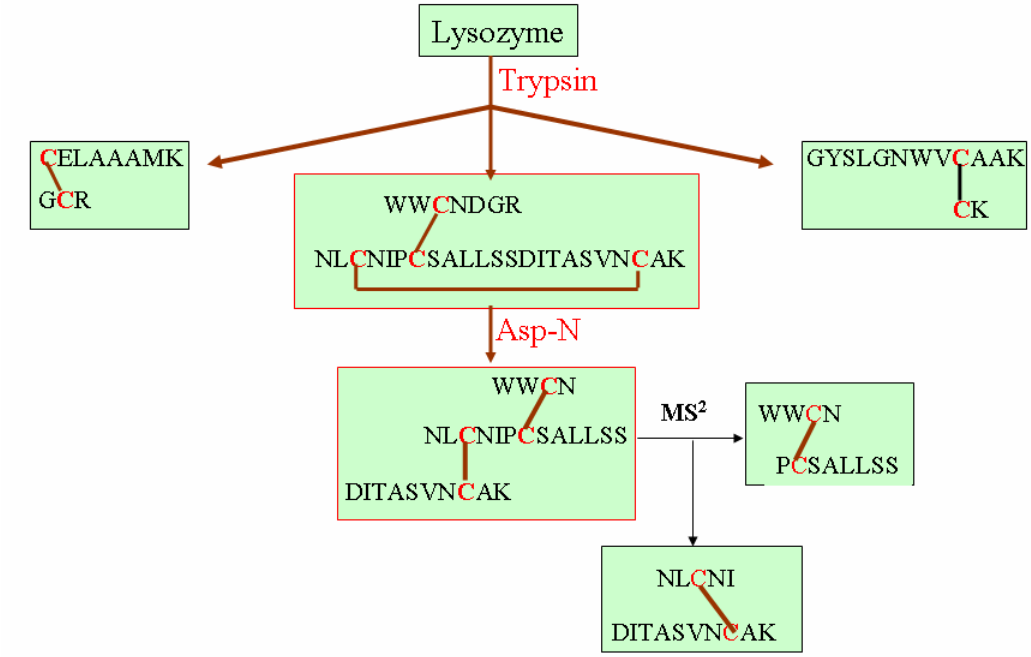

Fig. S13: Scheme of generation of the proteolytic peptides and the determination of disulfide connectivity in lysozyme.

Lysozyme:

Hen egg white lysozyme is a 129 residue protein that contains eight cysteines that form four disulfide bonds. There are 105 possible ways four disulfide bonds can be formed between eight cysteine residues. The intact native protein was subjected to trypsin digestion and subsequently analyzed through LC-MS/MS analysis. MS profile of the tryptic digest shows 8 peptides, for which the structures were obtained through the *DisConnect* (Fig. S10). While the peptide [**1**]-[**5**] corresponds to peptides without cysteines; peptide [**6**]-[**8**] contain cysteines. A quick inspection of these peptide structures reveals that both peptide [**6**] and [**7**] contain a pair of cysteine each. This unambiguously proves that the pair of cysteine present in each of these peptides are disulfide bonded, determining the connectivity for two of the four disulfide bonds. To find the connectivity of the other two disulfide bonds, we turn our analysis to peptide [**8**] that contains the remaining two disulfide bonds. Presence of two Asp residues in

this peptide presents a possibility of further proteolytic digestion by Asp-N, which results in a shorter proteolytic peptide (Peptide [9], Fig. S11(a)). Fig. S11 and S12 show the CID fragmentation spectra of $(M+3H)^{3+}$ species of [**9**] and the structures of the corresponding fragment ions. The structure of singly charged ion at m/z 1382.69 unambiguously brings out the connectivity pattern. The structure contains a Cys residue in each of the segment (C1 and C3), thus dictating a C1-C3 connectivity. Similarly the structure of the ion at m/z 788.91^{2+} establishes the C2-C4 connectivity. This means that the cysteine connectivity in [**9**] is C1-C3/ C2-C4. Interestingly, the isotopic distribution of this ion merges with the same of a doubly charge ion of m/z 1383.60. Expansion of the m/z region establishes the existence of two ions. This connectivity is further supported by the both probable structures of the ion at m/z 1190.61. Both of these two structures demand a disulfide bond between C1 and C3. This also illustrates the point that it is not essential to determine structure of every fragment ion with absolutely certainty to determine the disulfide connectivity. This proves the overall disulfide connectivity in lysozyme to be Cys1-Cys8, Cys2-Cys7, Cys3-Cys5, and Cys4-Cys6 (the Cys residues are number as Cys_i, where i stands for the residue number as per the overall protein sequence, analysis schematically summarized in Fig. S13).

SI Materials and Method

A. Experimental Protocol

Protocol of synthesis of foldamers

The linear peptide is synthesized using standard Fmoc chemistry, as described previously. The folding reaction was carried by dissolving 5 nmol of linear peptide in 200μL of oxidation buffer containing 100mM NH_4HCO_3 (pH 8.0) and 10% of dimethyl sulfoxide (DMSO). The progress of reaction was monitored using mass spectrometry and quenched after 24 hr by acidification with formic acid (10% final concentration). Reaction mixture was subjected to C_{18} analytical column, peptides were eluted over a linear gradient of 20-26 % acetonitrile and fractions were detected at 226 nm.

B. Description of input/output format of *DisConnect*

I. Analyses of proteolytic fragments:

Input/Output Format

The entire polypeptide sequence, experimental m/z values, along with their charge states, and a choice of protease are the primary input. In the case of multiple peptide chains, an X is inserted between every two peptide chains. Paste the protein sequence, ending with a * symbol, in the file "prot_seq". Put the m/z values and the charge states of the queried ions, in the format 'm/z value'space'charge state' (e.g. 1020.2 3, where the m/z value is 1020.2 and charge state 3) in the file 'peak_mass_ms'. Depending on the resolution of the experimental setup, the mass error range and mass type (monoisotopic / average) can be chosen. The user also has a handle on choosing the number of miscleavages. The theoretical peptide structures, with their masses within the user specified error range from the queried values, are the outputs. Their corresponding (M+H)+ values and the number of Cys residues present are also shown. . In the output for probable structures, discontinuous peptide chains (hereafter referred to as segments) are separated by a '|'. For chemical feasibility, it is imperative that such segments must be held together by S-S bonds. A provision is also made in *DisConnect* to study the structure of ions arising through probable neutral losses. This option is

useful when a key product ion is not obtained, but an ion obtained by subsequent neutral loss is present.

II. Analysis of the MS^n fragment ions:

Input/Output Format

The sequence of the polypeptide is entered by the user. Polypeptide sequence can be entered using the standard one letter codes. For peptide containing multiple chains, like insulin, a letter X should be entered between the two chains (e.g if a peptide contains two polypeptide chains with sequence GVCSF and RLTCY then the input is GVCSFXRLTCY). For user benefit, if these peptides are results of proteolytic digestion then an input file in the Result_MS folder is created, named "inp_MSn_match_MS_protein name.out" that contains the peptide sequence in the required format for the MS2 analysis. The user can copy the respective sequence from there and paste it. For MSn analysis, format of the complete sequence is as above. For the daughter ion (the ion undergoing MSn fragmentation) the input sequence is given inside the Resut_MSn folder, termed as input_for_MSn_rigorous/smart_Entered complete/fragment sequence. Copy the corresponding structure of the ion undergoing MSn fragmentation from this file. The m/z of the MSn fragment ions, with the charge states, also goes into the program as user input. Put the m/z values of the fragment ions, in the format 'm/z value'space'charge state' (e.g. 1020.2 3, where the m/z value is 1020.2 and charge state 3)in the file 'peak_mass_ms2' (for MS2) or 'peak_mass_msn' (for MSn). Depending on the resolution of the experimental setup, user has the freedom to tune the error range in mass accuracy and mass type (monoisotopic/average). An array of other input choices is available to the user. The output of the program contains probable structure/s of each MSn fragment ion. In the output for probable structures, discontinuous peptide chains (hereafter referred to as segments) are separated by a ‘|’. For chemical feasibility, it is imperative that such segments must be held together by S-S bonds. For those Cys containing outputs that have iterative Cys residues (number of Cys > 2[n-1], n being the number of segments), the possible residue mass arrangements of the Cys are also shown. It is to be noted that the output shows the calculated m/z values within the user specified mass error range. In this present study,

performed in ion trap mass spectrometer, we have used a lenient m/z cut off of 0.2Da while querying both the proteolytic and MSn fragments. All the program outputs, shown, are derived using the 'smart' mode of *DisConnect*.

References:

(1) H. R. Morris and P. Pucci, *Biophys. Biochem. Res. Commun.* 1985*,* **126***,* 1122-1128.

(2)(a). R. Yazdanparast, P. Andrews, D. L Smith and J. E. Dixon, *Anal. Biochem.* 1986*,* **153***,* 348-353. (b) R. Yazdanparast, P. Andrews, D. L Smith and J. E. Dixon, *J. Biol. Chem*., 1987, **262**, 2507-2513.

(3) C. Caporale, C. Sepe, Caruso, P. Pucci and V. Buonocore, *FEBS Lett.,* 1996, **393***,* 241-247.

(4) D. Fenyo, *Bioinformatics* 1997, **13***,* 617-618

(5) R. Craig, O. Krokhin, J. Wilkins and R. C. Beavis, *J. Proteome Res.,* 2003, **2***,* 657-661.

(6) P.R. Baker and K.R. Clauser, http://prospector.ucsf.edu

(7) W. Murad, R. Singh and T-Y. Yen, *BMC Bioinf.*, 2011, 12, (Suppl 1):S12

(8) K. Gupta, M. Kumar and Balaram, P. *Anal. Chem.,* 2010*,* **82***,* 8313-8319.