Published in final edited form as:
J Proteome Res. 2011 February 4; 10(2): 578–591. doi:10.1021/pr100764a.

Analysis of the disulfide bond arrangement of the HIV envelope protein CON-S gp140 ΔCFI shows variability in the V1 and V2 regions

Eden P. Go[1], **Ying Zhang**[1], **Sushma Menon**[1], and **Heather Desaire**[1,*]
[1]Department of Chemistry, University of Kansas, Lawrence KS 66047

Abstract

Disulfide bonding of cysteines is one of the most important protein modifications, and it plays a key role in establishing/maintaining protein structures in biologically active forms. Therefore, the determination of disulfide bond arrangement is one important aspect to understanding the chemical structure of a protein and defining its functional domains. Herein, aiming to understand how the HIV-1 envelope protein's structure influences its immunogenicity, we used an MS-based approach, liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance (LC/ESI-FTICR) mass spectrometry, to determine the disulfide linkages on an oligomeric form of the group M consensus HIV-1 envelope protein (Env), CON-S gp140 ΔCFI. This protein has marked improvement in its immunogenicity, compared to monomeric gp120 and wild-type forms of gp140 Envs. Our results demonstrate that the disulfide connectivity in the N-terminal region of CON-S gp140 ΔCFI is different from the disulfide bonding previously reported in the monomeric form of gp120 HIV-1 Env. Additionally, heterogeneity of the disulfide bonding was detected in this region. These data suggest that the V1/V2 region does not have a single, conserved disulfide bonding pattern, and that variability could impact immunogenicity of expressed Envs.

Keywords

disulfide; mass spectrometry; HIV-1; envelope protein; LC/ESI-FTICR MS

Introduction

Disulfide connectivity is one of the most essential protein modifications, and it plays a key role in stabilizing protein structures.[1–4] Disulfide bonds are usually formed between thiol groups of two cysteine residues that are proximal in three-dimensional structures in the secreted and non-cytoplasmic proteins.[4, 5] The formed disulfide bonds provide a significant reinforcement to the native folding state of the protein by decreasing the conformational entropy of the unfolded state, thus, stabilizing the protein.[4] Numerous studies have shown that correct disulfide connectivity in proteins is important for protein folding and becomes an indispensable criterion to assess the quality of protein in recombinant DNA technology.[3, 5–7] Given that disulfide linkages are essential in establishing and maintaining the proper

*To whom correspondence should be addressed: Department of Chemistry, 1251 Wescoe Hall Drive, University of Kansas Lawrence, KS, 66045 Phone: 785-864-3015 hdesaire@ku.edu.

Supporting Information
Tandem mass spectra of disulfide-linked peptides in CON-S gp140 ΔCFI. This material is available free of charge at http://pub.acs.org.

protein folds, the determination of disulfide bond arrangement is an important aspect to understanding protein chemical structures and defining the functional domains of proteins.[7–9]

If an x-ray crystal structure of a protein is available, the disulfide bonds can be determined from the protein structure directly.[10, 11] Likewise, if a protein structure has been assigned by nuclear magnetic resonance (NMR), the disulfide bonds are also determinable in a straightforward fashion.[12] However, if this level of structural data is not available *a priori*, other methods are more appropriate. For example, several analytical techniques, including high performance liquid chromatography (HPLC),[2, 13, 14] paper electrophoresis,[15] and Edman sequencing[2, 16] have been employed in the determination of disulfide bonds in proteins. These techniques were used to originally characterize the disulfide bonding network in a monomeric HIV-1 gp120 envelope protein (Env), from the IIIB isolate of HIV-1 expressed in Chinese Hamster Ovary (CHO) cells.[2] Typically, the disulfide bond determination by these techniques entails the enzymatic cleavage of proteins, followed by a separation process and then the detection of disulfide-linked peptides.[7] Although Edman sequencing is a powerful tool to obtain the amino acid sequence of disulfide-linked peptides, it requires that the proteolytic peptides are first completely separated from each other, generally by chromatography.[7, 17] As a result, this approach largely hinges on the ability to obtain highly purified peptides, when analyzing complex samples.[7] Mass spectrometry has emerged as an alternative analytical tool to Edman sequencing in determination of disulfide linkages in proteins, due to its ability to analyze peptide mixtures, along with its high sensitivity and throughput.

In addition to the analytical tools, disulfide mapping requires rigorous experimental controls, due to the possibility of disulfide rearrangement under certain circumstances. It has been previously observed that the disulfide bond scrambling could be triggered by the following conditions: (1) strong acid (e.g. 8.0 M sulfuric acid),[8, 9, 18] (2) the presence of a free thiol group on unpaired cysteine residues at neutral pH,[7–9, 18] and (3) the presence of a thiol group generated by hydrolytic cleavage of disulfide bonds during protein denaturation at neutral and alkaline pH. Thus, precautions must be taken when handling proteins to avoid disulfide bond rearrangement.

The focus of this paper is two-fold: to validate a protocol for profiling disulfide bonds on proteins, using liquid chromatography electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry (LC/ESI-FTICR MS) and to use the validated approach to analyze the disulfide bond arrangement in an oligomeric HIV-1 group M Consensus Env CON-S gp140 ΔCFI. CON-S gp140 ΔCFI was developed as a potential HIV vaccine candidate.[19] It is a synthetic form of Env representing the group M consensus Envs, and it induces neutralizing antibodies against type one and limited type two strains of the HIV-1 virus.[19] It also shows improved immunogenicity, compared to wild-type Envs.[19]

To date, the exact disulfide bonding patterns of oligomeric Env proteins, such as CONS gp140 ΔCFI have not been elucidated; however, the disulfide connectivity in several monomeric forms of HIV-1 Env, gp120, have been described.[2, 20–23] These previous studies indicate a conserved disulfide bonding pattern is present in various monomeric forms of HIV Envs, where all cysteine residues are utilized for disulfide connectivity. While these prior studies provide useful insight into the connectivity of monomeric Env, this form of the protein has been shown to perform inadequately as a vaccine candidate,[24,25] and many believe a key contributing reason is that a better mimic of viral Env would include protein that is present in an oligomeric form.[26–29] Several oligomeric HIV-1 Envs (dimers and trimers) have recently been expressed, and they are known to be covalently associated as a dimer or trimer through disulfide bonds.[19, 30] This surprising result implies that the

disulfide bonding network in recombinant oligomeric Env must be different from that in the previously characterized monomers, since at least one cysteine must be used for the intersubunit disulfide bond.

Previous studies have identified the location of interchain disulfide bonds between two subunits of Envs, by deleting certain domains and/or loops that are suspected to be involved in the connecting disulfide bond and detecting the changes in the number of domains that are covalently linked in the new protein.[30–32] Although deletion mutations on a protein are useful to identify which domains and/or loops are involved in the intersubunit contact of a protein, a possibility that truncating these segments influences the protein conformation and indirectly inhibits intersubunit association cannot be ruled out.[30–32] Additionally, these studies do not address the disulfide bonding which is not associated with the intersubunit linkage. Recent work by Finzi *et al.* indicates that dimeric Env is likely disulfide bonded in the V1/V2 region, although the full disulfide bonding network was not elucidated.[33]

The study presented herein is the first investigation to fully characterize the disulfide bonding network of the oligomeric Env, CON-S gp140 ΔCFI, using high resolution FTICR mass spectrometry. Our results show that two functional domains previously identified in a monomeric Env, the third and fourth variable regions (V3, V4), are also conserved in the disulfide bonding in CON-S gp140 ΔCFI. Among these two regions, V3 is the immunodominant region that is responsible for inducing type-specific neutralizing antibodies.[19] Additionally, heterogeneity in the disulfide bonding in the V1/V2 region of the protein is observed, and the expected disulfide bonding pattern in this part of the protein was not detected. In one case, two free cysteine residues were detected and these could potentially participate in forming oligomeric Env through interprotein disulfide bonds. Our results demonstrate that the oligomeric form of an HIV Env possess a different disulfide bonding network, compared to previously reported monomeric HIV Envs. These differences may be due to the fact that CON-S is a computationally derived recombinant protein, and other disulfide bonding studies have previously been carried out on wild-type recombinant proteins, implying that CON-S does not adopt a native disulfide bonding pattern. These results may also suggest that some variability exists in the possible disulfide bonding patterns of HIV-Envs, and if this variability does exist, it could potentially contribute to HIV's ability to evade the immune system.

Experimental Section

Materials and reagents

Albumin from bovine serum (BSA), 4-vinylpyridine, acetonitrile, formic acid, acetic acid, and ammonium bicarbonate were all obtained in high purity from Sigma-Aldrich (St. Louis, MO). The HIV envelope glycoprotein CON-S gp140 ΔCFI was expressed and purified in the laboratory of Dr. Barton F. Haynes, Duke Human Vaccine Research Institute (Duke University, Durham, NC), using the method described elsewhere[19]. Peptide *N*-glycosidase F (PNGase F) from *Elizabethkingia meningosepticum* and proteomics grade trypsin were purchased from Calbiochem (San Diego, CA) and Promega (Madison, MI), respectively. Ultra-pure water was obtained from a Millipore Direct-Q3 filtration system. (Billerica, MA).

Protein digestion

About 100 μg of either BSA or CON-S gp140 ΔCFI were alkylated with 4-vinylpyridine at a protein:vinylpyridine ratio of 1:6 (molar ratio) for 1 hour in the dark at room temperature to cap free cysteine residues prior to any enzymatic digestion. Following alkylation with 4-vinylpyridine, CON-S gp140 ΔCFI was deglycosylated with 1 μL of PNGase F enzyme solution (≥ 4500 units/mL) and was incubated at 37 °C for one week while BSA was digested with trypsin. In-solution trypsin digestion of both proteins was carried out at 37°C

for 18 hours at a 1:20 enzyme-to-protein ratio (w/w), followed by a second trypsin addition under the same conditions. The digestion was stopped by the addition of 1μL concentrated acetic acid and the resulting protein digest mixture was analyzed by reversed phase HPLC/ESI-FTICR MS.

LC/ESI-FTICR MS analysis

The tryptic digests were subjected to reversed phase high performance liquid chromatography (RP-HPLC, Dionex, Sunnyvale, CA) coupled with a hybrid linear ion trap Fourier transform ion cyclotron resonance (LTQ FTICR) mass spectrometer equipped with a 7 Tesla actively shielded magnet (ThermoElectron Corp, San Jose, CA). The mobile phase A was 99.9% water with 0.1% formic acid. Mobile phase B was 99.9% acetonitrile with 0.1% formic acid. Approximately 5 μL of the sample was injected and separated on a C18 PepMap 300™ column (150mm×300 μm i.d. 5 μM, 300Å; LC Packings, Sunnyvale, CA) at a flow rate of 5μL/min. The tryptic peptides were eluted using the following gradient, which was modified from the method described in the literature[34]: The mobile phase initially contained 2% B and the level of B increased linearly to 40% over 30 min, then ramped to 90% B over 20 min. The column was re-equilibrated after holding at 90% B for 10 min.

The samples were infused into the electrospray ion source, and the hybrid LTQ FTICR data acquisition was performed in a data-dependent scanning mode. Briefly, the MS^1 spectra were recorded in FT MS scan and the six most abundant peptide ions in the MS scan were sequentially selected for CID performed in the LTQ mass analyzer, with 35% normalized collision energy and a 3 min dynamic exclusion window. Electrospray ionization was achieved with a spray voltage of ~4.0 kV. Nitrogen was used as nebulizing gas and set at a pressure of 10 psi. The capillary temperature was maintained at 200 °C. All the data were acquired in the positive-ion mode and analyzed using Xcalibur 2.0 software (ThermoElectron Corp, San Jose, CA).

Results and Discussion

A disulfide-rich model protein, bovine serum albumin (BSA), was used to optimize experimental conditions and to validate that conditions used in this study do not cause disulfide scrambling. The optimized approach was subsequently applied in the determination of disulfide bond arrangement of HIV Env CON-S gp140 ΔCFI.

Disulfide mapping of bovine serum albumin

Bovine serum albumin (primary accession number P02769 from Swiss-Prot) contains ~600 amino acid residues, 77 potential cleavage sites for trypsin, and 35 cysteine residues.[35] The putative disulfide bond arrangements36 are shown in Figure 1, in which 17 inter- and intrachain disulfide linkages are formed. One unpaired cysteine residue, Cys^{58}, has a free thiol group, and it could promote the disulfide bond rearrangement at neutral pH ~7.0.[7] To preclude disulfide scrambling, the thiol groups in the unpaired cysteine residues were capped with the addition of excess alkylating reagent, 4-vinylpyridine. 4-Vinylpyridine was used as the alkylating agent because the modified peptides containing *S*-β-(4-pyridylethyl)-cysteine can be readily identified, due to the presence of characteristic ion, *S*-pyridylethyl moiety, upon CID experiments in mass spectrometry.[37] The protein was then digested with trypsin at neutral pH ~7.0. Trypsin was used for generating the disulfide-linked peptides due to its substrate specificity and its capability to produce relatively simple mass spectra, compared to other proteases (e.g. pepsin).[7] The resultant tryptic peptides containing disulfide linkages and modified unpaired cysteine residues were then subjected to reversed phase HPLC followed by ESI-FTICR MS and MS/MS analysis, as shown in Figure 2.

When the tryptic digestion mixtures were separated, several disulfide-linked peptides with inter- or/and intrachain disulfide linkages were identified. A representative reversed phase HPLC chromatogram and LC/ESI-FTICR mass spectrum of tryptic peptides from BSA are shown in Figure 3A and B, respectively. The FTICR MS mass spectrum of the HPLC fraction eluting between 16–17 min (the red bar shown in Figure 3A) shows two peaks at *m/z* 925.8946 and *m/z* 1234.1892 with different charge states of the same disulfide-linked peptide species. Figure 3 inset shows the zoom-in spectra used to identify the monoisotopic peaks at *m/z* 925.3933, *m/z* 1234.5219 and their corresponding charge states (+4, and +3, respectively). By comparing the observed masses with the theoretically calculated masses of the putative disulfide-linked peptides, the peaks at *m/z* 925.3933 and *m/z* 1234.1892 correspond to the 3+ and 4+ charge state of the disulfide-linked three peptide chain found in BSA. Table 1 summarizes all the experimentally determined disulfide linkages in BSA. All the 35 cysteines have been identified by LC/ESI-FTICR MS analysis. Each group of peptides containing disulfide linkages or peptides with unpaired cysteine residues has been identified with high mass accuracy and confirmed by MS/MS, and was consistent with the known disulfide linkages in BSA, as described in literature.[15,36]

To further verify the disulfide linkages and peptide sequences of the disulfide-linked peptides or peptides with unpaired cysteine residues, MS/MS analyses were performed using data dependent acquisition in the linear ion trap. Representative ESI-MS/MS data are shown in Figure 4. Three classes of peptides containing disulfide-linked or unpaired cysteine residues have been identified by the product ions observed in MS/MS analysis. Figure 4A depicts the CID data of a peptide that contains a cysteine capped with 4-vinylpyridine. The major ions in the spectrum are consistent with the expected peptide sequence that retains the chemically modified thiol group on the cysteine. Figure 4B represents an intrachain disulfide-linked peptide from BSA. For the product ions of *m/z* 438.09 and 566.39, in which the disulfide bonds are dissociated, the cysteine residues were detected as free thiol groups (−SH), as expected.[4] Again, the MS/MS data are consistent with the depicted peptide sequence. Figure 4C demonstrates the identification of a three peptide chain disulfide-linked peptide. The product ions denoted by $^{\alpha}y/^{\alpha}b$ were generated from the peptide chain containing an N-terminal serine (S), whereas product ions denoted by $^{\beta}y/^{\beta}b$ and $^{\gamma}y/^{\gamma}b$ were produced from the peptide chains with N terminal glutamic acid (E) and asparagine (N), respectively. As demonstrated here, in addition to using high mass accuracy data in FTICR MS analysis to identify the disulfide-linked peptides in BSA, the peptide sequences and disulfide linkages in these peptides were also confirmed with MS/MS experiments, which significantly increases the confidence level in the mass assignments and implies that this approach is an effective way to determine the disulfide bonds in complex protein digests.

Disulfide mapping of CON-S gp140 ΔCFI

Owing to the successful application of this approach in mapping the disulfide bonds of bovine serum albumin, we extend the utility of this approach to a structurally complex protein, CON-S gp140 ΔCFI. CON-S gp140 ΔCFI is a computationally generated group M consensus form of the HIV-1 Env with shorter variable loops (V1–V5) compared to wild-type variable loop sequences[19]; that when used as an immunogen in small animal vaccine trials, was shown to induce type I and limited type II neutralizing antibodies against the HIV-1 virus.19 This protein contains ~600 amino acid residues19, 52 potential cleavage sites for trypsin, 18 cysteine residues, and 31 potential N-linked glycosylation sites that are appended with a heterogeneous population of glycans.38 It is expressed as a mixture of monomers, dimers, and trimers. According to the disulfide connectivity in a previously analyzed recombinant, monomeric HIV envelope protein, gp120,[2] nine disulfide bonds should be present for each monomeric unit of CON-S gp140 ΔCFI, if the expected disulfide bonding pattern is conserved. This expected disulfide connectivity for the monomeric form

of CON-S gp140 ΔCFI is shown in Figure 5A, in which the disulfide bonds define the boundaries of the variable regions (V1–V5) as well as conserved domains in HIV Env protein.[2, 39] Since CON-S gp140 ΔCFI is an oligomeric form of Env, we accounted for the possibility that the disulfide bonding pattern could be significantly different. Therefore, many possibilities for the disulfide connectivity were considered and investigated in data analysis, including free cysteine residues, the disulfide bonding pattern observed previously, 2 new disulfide bonds within the monomeric unit and interchain disulfide bonds between two monomeric units.

One final important experimental consideration in analyzing disulfide bonds of CON-S gp140 ΔCFI is dealing with protein glycosylation. This protein is extensively glycosylated. The glycans are bulky and block proteases from accessing the peptide backbone. This would result in multiple missed cleavage sites during enzymatic digestion, thus producing complicated spectra in MS analysis. As a result, the glycans attached to CON-S gp140 ΔCFI were cleaved with PNGase F prior to the disulfide mapping, to facilitate trypsin digestion and data analysis.

Control experiment to validate the deglycosylation step in disulfide mapping of CON-S gp140 ΔCFI

While the deglycosylation step is essential in assigning the disulfide bond arrangement of CON-S gp140 ΔCFI, it is possible that this process would result in disulfide scrambling. Herein, we introduced a control experiment to investigate this possibility. The validation protein, bovine serum albumin, was alkylated with 4-vinylpyridine and was subsequently incubated with PNGase F at 37 °C for one week, then digested with trypsin. The resulting peptide mixture containing inter-/intradisulfide linkages were subjected to LC/ESI-FTICR MS analysis. This study showed that the same disulfide connectivity in BSA was obtained, compared to the results described in the previous section, demonstrating that the conditions described herein for the extra deglycosylation step do not affect the determination of disulfide linkages in glycoproteins.

Determination of disulfide bonds in CON-S gp140 ΔCFI

A schematic workflow of the disulfide mapping experiment for CON-S gp140 ΔCFI is shown in Figure 2. The glycoprotein was alkylated with 4-vinylpyridine and was enzymatically deglycosylated with PNGase F. The enzymatic release of glycans converts asparagine in the consensus sequence, NXT/S, to aspartic acid, resulting in a mass shift of 1 Da; as shown in Scheme 1.

It is necessary to know which asparagine (N) would be converted to aspartic acid (D) during deglycosylation, so that the theoretical masses of each peptide can be elucidated; therefore, the glycosylation site occupancy of the protein must be known or determined. The determination of glycosylation site occupancy in CON-S gp140 ΔCFI was achieved and described in a previous work.[38] Accordingly, the cysteine-containing tryptic peptides from CON-S gp140 ΔCFI with fully or partially occupied glycosylation sites and their corresponding deglycosylated peptide masses can be deduced and are listed in Table 2. The masses in this table were used to generate theoretical masses for a wide variety of possible disulfide-linked peptides, as described above, and this list of theoretical masses was interrogated against the mass spectral data, to determine the disulfide connectivity of this protein.

The mass spectral data were acquired by first subjecting the deglycosylated CON-S gp140 ΔCFI to tryptic digest, then obtaining an LC/ESI-FTICR MS analysis. A representative reversed phase HPLC chromatogram and ESI-FTICR mass spectrum of tryptic peptides

from deglycosylated CON-S gp140 ΔCFI are shown in Figure 6A and Figure 6B, respectively. The FTICR MS mass spectrum of the HPLC fraction eluting between 14–15 min (the red bar shown in Figure 6A) shows a doubly charged peak at *m/z* 788.8779 originating from an intrachain disulfide linked peptide, LINCDTSAITQACPK. Figure 6 inset shows a zoom-in spectrum of this peak, which is used to identify the monoisotopic peak at *m/z* 788.3805 and its corresponding charge state. The aspartic acid residue (labeled in green) was generated upon deglycosylation with PNGase F, as described above.

The summary of all the experimentally determined disulfide linkages in CON-S gp140 ΔCFI is shown in Table 3. These assignments are confirmed by MS/MS experiments (see Supplementary Data). One striking feature of these data is that several of the disulfide-bonded peptides were detected in multiple forms. For example, two peptides were detected as being capped with 4-vinylpyridine and also detected as being bonded to each other. Also, the two peptides, LINCDTSAITQACPK and LTPLCVTLDCTNVDVTNTTNNTEEK were detected as being bound to each other (to form a “box” with two disulfide bonds connecting them) as well as detected individually, as internally linked species. One additional feature to note is that for the C-terminal end of the protein, the disulfide bonding pattern was consistent with the disulfide bonding observed in the previous analyses of Envs.[2] This region specifically encapsulates the third and fourth variable regions (V3, V4) in CON-S gp140 ΔCFI. It is worth noting that the V3 loop is an immunodominant region, which is responsible for the generation of V3 targeted neutralizing antibodies.

To further verify each of the disulfide linkages and peptide sequences of these peptide clusters, MS/MS analysis was performed. In each case, the MS/MS data are consistent with the assignments presented in Table 3. Representative LC/ESI-MS/MS data are shown in Figure 7. Figure 7A depicts the fragmentation of a peptide with an intrachain disulfide linkage. The ions at *m/z* 975, 1106, and 1203, all contain an intact disulfide bond. Other significant ions are typical y and b fragments. All of the other major ions in the spectrum are also assignable and consistent with the peptide sequence shown in the Figure.

Figure 7B represents an example verification of an interchain disulfide-linked peptide. In Figure 7B, the product ions denoted by $^{\alpha}y/^{\alpha}b$ were generated from the peptide chain with an N-terminal threonine (T), whereas product ions denoted by $^{\beta}y/^{\beta}b$ were produced from the peptide chains with N-terminal isoleucine (I). As in Figure 7A, the data in 7B clearly show that the MS/MS data is highly consistent with the assignment for each precursor ion, further validating the data in Table 3.

An overview of the experimentally determined disulfide bonds from this study is shown in Figure 5B. In the N-terminal region (defined by amino acid residues 1–210) of CON-S gp140 ΔCFI, substantial differences were detected, compared to the expected disulfide bonding configuration. Heterogeneity was present, and this is indicated in the figure with burgundy bonds connecting the cysteine residues in alternate configurations, compared to the yellow-orange bonds. Also notable is that *neither* disulfide bonding pattern in the V1/V2 region of the analyzed protein matches data of the previously characterized Env disulfide bond (shown in 5A). The differences in the connectivity of the V1/V2 region could clearly impact the epitope accessibility of the protein; a pictorial representation of the different disulfide connectivities detected is shown in Figure 8. While the disulfide connectivity in the N-terminal region of the protein is substantially different from the expected disulfide connectivity, the disulfide linkages from the V3 and V4 regions are consistent with the putative disulfide bonds in a previous analysis of a monomeric form of Env.[2]

As it is known that oligomeric forms of Env, including this one, can form their oligomeric species through disulfide bonding, the data here suggest that the cysteine residues

responsible for this are in the V1/V2 region of the protein. These data are consistent with other researchers who have assigned this general region as being responsible for interprotein disulfide bonding in recombinant Env preparations.[33] While no interprotein disulfide bonds were specifically verified in our work, the data suggest that likely candidate sites include the cysteine residues, C125, C130, C200, and C209, that were detected as having multiple possible binding partners as shown in Figure 8. Clearly a single disulfide bonding network is not strongly favored in the V1/V2 region of the protein, and it seems feasible that one or several of the cysteines in this region could participate in forming alternative connection patterns, such as intersubunit disulfide bonding.

Biological implications of this study

The results obtained from this study suggest that the disulfide connectivity of CON-S gp140 ΔCFI is different from the disulfide linkages in the previously analyzed monomeric form of HIV-1 Env expressed from CHO cells[2] (Figure 5A), especially in the N-terminal region of protein where the first and second variable regions (V1, V2) are normally formed. One possible explanation for variability in disulfide bonding is that CON-S is a synthetic *Env*, which has shorter variable loops[19] and lower infectivity compared to the wild-type Env that was previously analyzed. While the study herein provides the first *direct evidence* of disulfide variability in the V1 and V2 regions of an Env, other researchers have observed *indirect evidence* of disulfide variability in a variety of other wild-type Envs.

Our findings complement a previous study by Jobes *et al*, which shows that the V1 and V2 regions of Env are highly variable, with the respect to the cysteine content.[39] These authors suggest that when the insertion or deletion of cysteine residues occurs in these regions, it could result in the disruption of normal disulfide bonds.[39] Both our work and this previous work show that the cysteine residues in the N-terminal region of Env can have a varying disulfide connectivity. The V1 and V2 region of the protein are important, because they shield the V3 region, preventing antibodies from directly accessing neutralizing epitopes present in V3.[39–41] Since the disulfide linkages of the V1 region in CON-S gp140 ΔCFI are connected differently than those described previously,[2, 20–23] and two unpaired cysteine residues are present, the disulfide connectivity in CON-S gp140 ΔCFI may cause the epitopes in the V3 region to be more exposed, facilitating immune response to this region.

The fact that an unexpected disulfide profile existed in the V2 region of the protein, for this oligomeric Env, is also significant in the context of Walker *et. al.'s* work[42] and the work of Haynes, Tsao, and Bonsignori.[43] These researchers have recently identified V2 as a key region that elicits the broadly neutralizing antibodies PG9[42] PG16,[42] and CH01.[43] The quaternary antibodies PG9, PG16, and CH01, all preferentially react with trimers in native conformations on virions and generally do not react with recombinant envelope proteins.[42, 43] One possible explanation for this finding is that the proteins' disulfide bonding is not the same on the trimeric species as it is typically generated in monomeric Env. Interestingly, Haynes Tsao and Bonsignori at Duke have found transmitted/founder recombinant gp140 envelope proteins that do bind quaternary-epitope antibodies.[43] It will be of interest to perform disulfide bond analysis on those rare recombinant envelopes that do bind quaternary epitope-reactive monoclonal antibodies to determine a putative native disulfide bonding pattern.

Aside from the disulfide linkages in V1 and V2 regions, the final interesting implication of this work relates to the disulfide linkage in the V3 and V4 loop, which is identified in our study to be consistent with the disulfide bonding established for monomeric Env.[2, 20–23] This is somewhat expected, since this region was determined to be immunodominant, and it contributes to the improved immunogenicity in CON-S gp140 ΔCFI[19] Certainly, the protein would need to be folded properly in the V3 and V4 regions, if it were to elicit V3-targeted

antibodies against numerous strains of HIV-1, which it does.[19] The question that remains, however, is whether the unique disulfide arrangement in the earlier part of the protein *contributed to* CON-S's improved immunogenicity or *detracted from* it.

Conclusions

We implemented an MS-based approach to characterize the disulfide linkages on the HIV-1 Env, CON-S gp140 ΔCFI. Prior to applying this approach to the Env, the method was validated using a disulfide-rich protein, bovine serum albumin. In this protein, all the 35 cysteine residues were detected. A total 17 disulfide-linked peptides and one unpaired cysteine residue were identified with high mass accuracy, and confirmed by MS/MS experiments. This data is in agreement with the putative disulfide linkages reported previously for BSA.[36]

This approach was then applied to elucidate the disulfide connectivity in recombinant HIV-1 Env, CON-S gp140 ΔCFI. All of the 18 cysteine residues were identified, and some were detected in multiple forms. In particular, heterogeneity existed in the N-terminal region of CON-S gp140 ΔCFI,, and the identified disulfide bonds did not match the bonding pattern that has been previously published. These findings suggest that the N-terminal region of Env does not have a highly conserved disulfide bonding pattern. Although the cysteine residues in the N terminal region of the protein are different than the putative disulfide linkages from an analysis of a monomeric form of HIV-1 envelope protein,[2] the immunodominant V3 loop and V4 loop are indeed identified in CON-S gp140 ΔCFI.

Overall, the MS-based approach employed in this study has proven to be useful in the determination of disulfide bonding patterns of BSA and a recombinant HIV-1 Env, CON-S gp140 ΔCFI and should be applicable to other proteins. While the MS analysis of the disulfide-linked peptides generated from the tryptic digest of these proteins is straightforward, the task of analyzing the MS/MS data to unambiguously identify the disulfide-linked peptides is painstaking and time consuming, especially for disulfide-rich proteins. This mainly stems from the lack of suitable software tools and/or web-based programs that would automate the analysis. With the recent advances in proteomics methodologies and the introduction of new hybrid mass spectrometers, new algorithms are emerging to meet new demands for suitable informatics tools to address this issue.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge the National Institutes of Health for funding (project number RO1GM077266). The authors also thank Dr. Hua-Xin Liao at Duke University Medical Center, for providing the recombinant CON-S gp140 protein, and Dr. Todd Williams at the University of Kansas Applied Proteomics Laboratory for granting access to the FTICR MS instrument.

References

1. Fukuyama Y, Iwamoto S, Tanaka KJ. Rapid sequencing and disulfide mapping of peptides containing disulfide bonds by using 1,5-diminonaphthalene as a reductive matrix. J. Mass. Spectrom. 2006; 41:191–201. [PubMed: 16382486]
2. Leonard CK, Spellman MW, Riddle L, Harris RJ, Thomas JN, Gregory TJ. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. 1990; 265:10373–10382. [PubMed: 2355006]

3. Wefing S, Schnaible V, Hoffmann D. SearchXLinks. A program for the identification of disulfide bonds in proteins from mass spectra. Anal. Chem. 2006; 78:1235–1241. [PubMed: 16478117]
4. Zhang M, Kaltashov IA. Mapping of protein disulfide bonds using negative ion fragmentation with a broadband precursor selection. Anal. Chem. 2006; 78:4820–4829. [PubMed: 16841900]
5. Nguyen DN, Bechker GW, Riggin RM. Protein mass spectrometry: application to analytical biotechnology. J. Chromatogr. A. 1995; 705:21–45. [PubMed: 7620570]
6. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in *Escherichia Coli*. Nat. Biotechnol. 2004; 22:1399–1408. [PubMed: 15529165]
7. Gorman JJ, Wallis TP, Pitt JJ. Protein disulfide bond determination by mass spectrometry. Mass Spectrom. Rev. 2002; 21:183–216. [PubMed: 12476442]
8. Smith DL, Zhou Z. Strategies for locating disulfide bonds in proteins. Methods Enzymol. 1990; 193:374–389. [PubMed: 2074827]
9. Haniu M, Arakawa T. Analysis of disulfide structures in proteins. Curr. Top. Pept. Protein Res. 1997; 2:115–124.
10. Craik DJ, Daly NL. NMR as a tool for elucidating the structures of circular and knotted proteins. Mol. Biosyst. 2007; 3:257–265. [PubMed: 17372654]
11. Acharya KR, Lloyd MD. The advantages and limitations of protein crystal structures. Trends Pharmacol. Sci. 2005; 26:10–14. [PubMed: 15629199]
12. Liu H, Hsu J. Recent developments in structural proteomics for protein structure determination. Proteomics. 2005; 5:2056–2068. [PubMed: 15846841]
13. Andrews PC. Selective isolation of disulfide-containing peptides from trypsin digests using strong cation exchange HPLC. Curr. Res. Protein Chem. 1990:95–102.
14. Crimmins DL. Analysis of disulfide-linked homo- and hetero-peptide dimmers with a strong cation-exchange sulfoethyl aspartamide column. Pept. Res. 1989; 2:395–401. [PubMed: 2520779]
15. Brown JR, Kauffman DL, Hartley BS. Primary structure of porcine pancreatic elastase; The N-terminus and disulfide bridges. Biochem. J. 1967; 103:497–507. [PubMed: 5340368]
16. Haniu M, Arakawa T, Bures EJ, Young Y, Hui JO, Rohde MF, Welcher AA, Horan T. Human leptin receptor. Determination of disulfide structure and N-glycosylation sites of the extracellular domain. J. Biol. Chem. 1998; 273:28691–28699. [PubMed: 9786864]
17. Irungu J, Go EP, Dalpathado DS, Desaire H. Simplification of mass spectral analysis of acidic glycopeptides using GlycoPep ID. Anal. Chem. 2007; 79:3065–3074. [PubMed: 17348632]
18. Owusu-Apenten RK, Chee C, Hwee OP. Evaluation of a sulphydryl-disulphide exchange index (SEI) for whey proteins — β-lactoglobulin and bovine serum albumin. Food Chem. 2003; 83:541–545.
19. Liao H, Sutherland LL, Xia S, Brock ME, Scearce RM, Vanleeuwen S, Alam SM, McAdams M, Weaver EA, Camacho ZT, Ma B, Li Y, Decker JM, Nabel GJ, Montefiori DC, Hahn BH, Korber BT, Gao F, Haynes BF. A group M consensus envelope glycoprotein induces antibodies that neutralize subsets of subtype B and C HIV-1 primary viruses. Virology. 2006; 353:268–282. [PubMed: 17039602]
20. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998; 393:648–659. [PubMed: 9641677]
21. Wyatt R, Kwong PD, Desjardins E, Sweet RW, Robinson J, Hendrickson WA, Sodroski JG. The antigenic structure of the HIV gp120 envelope glycoprotein. Nature. 1998; 393:705–711. [PubMed: 9641684]
22. Zhou T, Xu L, Dey B, Hessell AJ, Van RD, Xiang S-H, Yang X, Zhang M-Y, Zwick MB, Arthos J, Burton DR, Dimitrov DS, Sodroski J, Wyatt R, Nabel GJ, Kwong PD. Structural definition of a conserved neutralization of epitope on HIV-1 gp120. Nature. 2007; 445:732–737. [PubMed: 17301785]
23. Chen B, Gong H, Skehel JJ, Wiley DC, Harrison SC. Structure of an unliganded simian immunodeficiency virus gp120 core. Nature. 2005; 433:834–841. [PubMed: 15729334]
24. Flynn NM, Forthal DN, Harro CD, Judson FN, Mayer KH, Para MF. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. J. Infect. Dis. 2005; 191:654–665. [PubMed: 15688278]

25. Gilbert PB, Peterson ML, Follmann D, Hudgens MG, Francis DP, Gurwith M, Heyward WL, Jobes DV, Popovic V, Self SG, Sinangil F, Burke D, Berman PW. Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. J. Infect. Dis. 2005; 191:666–677. [PubMed: 15688279]
26. Sharma VA, Kan E, Sun Y, Lian Y, Cisto J, Frasca V, Hilt S, Stamatotos L, Donnelly JJ, Ulmer JB, Barnett SW, Srivastava IK. Structural characteristic correlate with immune responses induced by HIV envelope glycoprotein vaccines. Virology. 2006; 352:131–144.
27. Srivastava IK, Stamatatos L, Legg H, Kan E, Fong A, Coates SR, Leung L, Wininger M, Donnelly JJ, Ulmer JB, Barnett SW. Purification and characterization of oligomeric envelope glycoprotein from a primary R5 subtype B human immunodeficiency virus. J. Virol. 2002; 76:2835–2847. [PubMed: 11861851]
28. Srivastava IK, Stamatatos L, Kan E, Vajdy M, Lian Y, Hilt S, Martin L, Vita C, Zhu P, Roux KH, Vojtech L, Montefiori DC, Donnelly J, Ulmer JB, Barnett SW. Purification, characterization, and immunogenicity of a soluble trimeric envelope protein containing a partial deletion of the V2 loop derived from SF162, an R5-tropic human immunodeficiency virus type 1 isolate. J. Virol. 2003; 77:11244–11259. [PubMed: 14512572]
29. Barnett SW, Srivastava IK, Ulmer JB, Donnelly JJ, Rappuoli R. Development of V2-deleted trimeric envelope vaccine candidates from human immunodeficiency virus type 1 (HIV-1) subtypes B and C. Microbes Infect. 2005; 7:1386–1391. [PubMed: 16275150]
30. Yuan W, Craig S, Yang X, Sodroski J. Inter-subunit disulfide bonds in soluble HIV-1 envelope glycoprotein trimers. Virology. 2005; 332:369–383. [PubMed: 15661168]
31. Center RJ, Earl PL, Lebowitz J, Schuck P, Moss B. The human immunodeficiency virus type 1 gp120 V2 domain mediates gp41-independent intersubunit contacts. J. Virol. 2000; 74:4448–4455. [PubMed: 10775580]
32. Center RJ, Leapman RD, Lebowitz J, Arthur LO, Earl PL, Moss B. Oligomeric structure of the human immunodeficiency virus type 1 envelope protein on the viron surface. J. Virol. 2002; 76:7863–7867. [PubMed: 12097599]
33. Finzi A, Pacheco B, Zeng X, Kwon YD, Kwon PD, Sodroski J. Conformational characterization of aberrant disulfide-linked HIV-1 gp120 dimers secreted from overespressing cells. J. Virol. Methods. 2010 In Press.
34. Ihling C, Berger K, Hoefliger MM, Fuehrer D, Beck-Sickinger AG, Sinz A. Nano-high-performance liquid chromatography in combination with nano-electrospray ionization Fourier transform ion-cyclotron resonance mass spectrometry for proteome analysis. Rapid. Commun. Mass Spectrom. 2003; 17:1240–1246. [PubMed: 12811746]
35. Schnaible V, Wefing S, Buecker A, Wolf-Kuemmeth S, Hoffmann D. Screening for disulfide bonds in proteins by MALDI in-source decay and LIFT-TOF/TOF-MS. Anal. Chem. 2002; 74:2386–2393. [PubMed: 12038765]
36. Brown JR. Structure of serum albumin: disulfide bridges. Fed. Proc. 1974; 33:1389–1389.
37. Moritz RL, Hall NE, Connolly LM, Simpson RJ. Determination of the disulfide structure and *N*-glycosylation sites of the extracellular domain of the human signal transducer gp130. J. Biol. Chem. 2001; 276:8244–8253. [PubMed: 11098061]
38. Go EP, Irungu J, Zhang Y, Dalpathado DS, Liao H-X, Sutherland LL, Alam SM, Haynes BF, Desaire H. Glycosylation site-specific analysis of HIV envelope proteins (JR-FL and CON-S) reveals major differences in glycosylation site occupancy, glycoform profiles, and antigenic epitopes' accessibility. J. Proteome Resch. 2008; 7:1660–1674.
39. Jobes DV, Daoust M, Nguyen V, Padua A, Michele S, Lock MD, Chen A, Sinangil F, Berman PW. High incidence of unusual cysteine variants in gp120 envelope proteins from early HIV type 1 infections from a phase 3 vaccine efficacy trial. AIDS. Res. Hum. Retroviruses. 2006; 22:1014–1021. [PubMed: 17067272]
40. Mckeating JA, Shotton C, Cordell J, Graham S, Balfe P, Sullivan N, Charles M, Page M, Bolmstedt A. Characterization of neutralizing monoclonal antibodies to linear and conformation-dependent epitopes within the first and second variable domains of human immunodeficiency virus type 1 gp120. J. Virol. 1993; 67:4932–4944. [PubMed: 7687306]

41. Pinter A, Honnen WJ, He Y, Gorny MK, Zolla-Pazner S, Kayman SC. The V1/V2 domain of gp120 is a global regulator of the sensitivity of primary human immunodeficiency virus type 1 isolates to neutralization by antibodies commonly induced upon infection. J. Virol. 2004; 78:5205–5215. [PubMed: 15113902]

42. Walker LM, Phogat SK, Chan-Hui P-Y, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR. Protocol G Principal Investigators. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science. 2009; 326:285–289. [PubMed: 19729618]

43. Haynes BF, Tsao CY, Bonisgnori M. Personal communication.

Figure 1.

Protein sequence and theoretical disulfide-linkage arrangement in bovine serum albumin; The signal peptide sequence is shown in blue, disulfide linkages are indicated with a yellow-oranges solid line, and cysteine residues are shown in red.

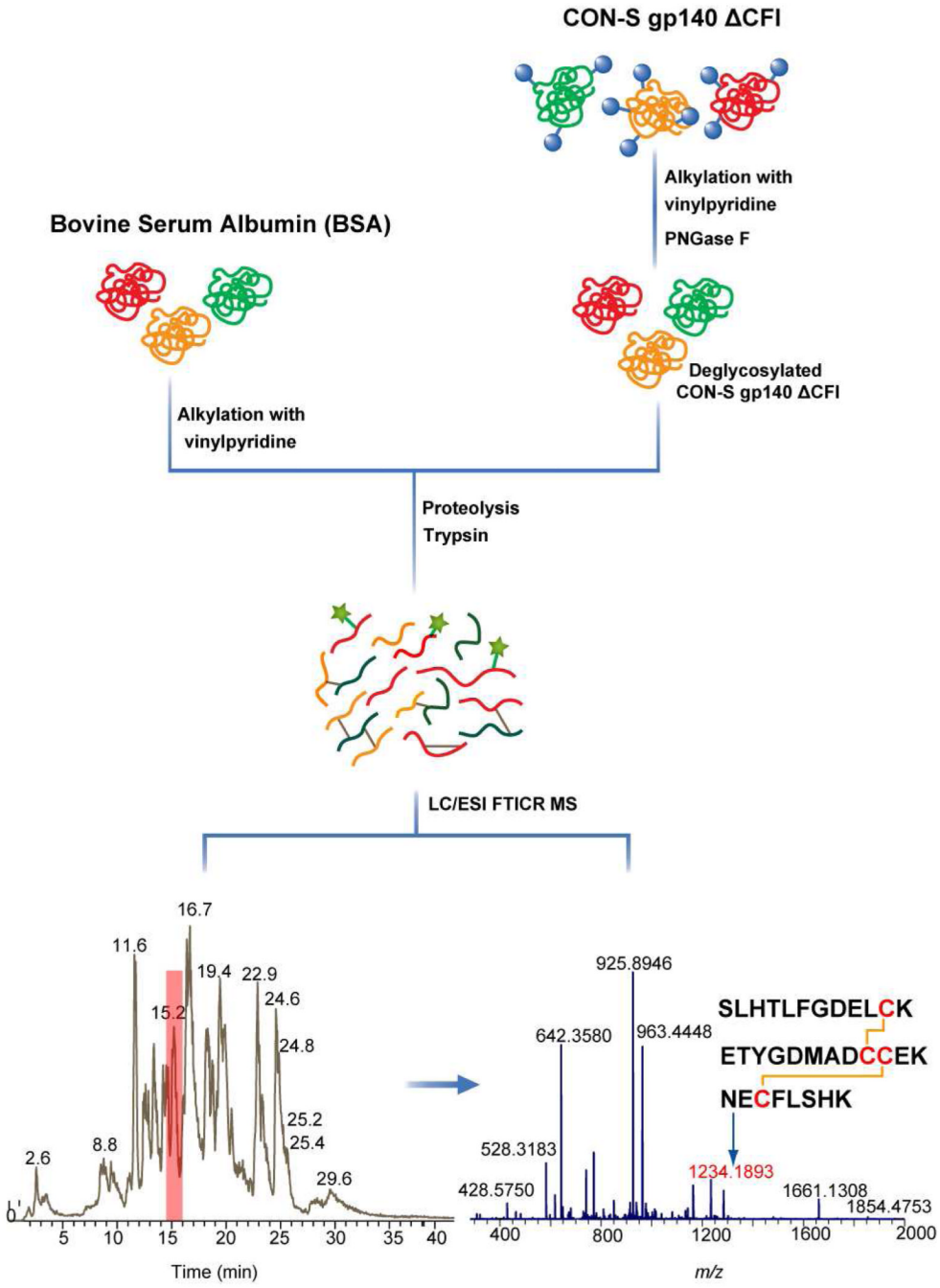


Figure 2.

Schematic representation of disulfide bond mapping workflow in bovine serum albumin and CON-S gp140 ΔCFI; The blue balls represent the glycans on CON-S gp140 ΔCFI. The yellow green stars indicate the free cysteines alkylated with 4-vinylpyridine.

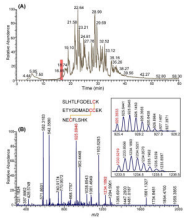


Figure 3.
Representative (A) reversed phase HPLC chromatogram and (B) LC/ESI-FTICR MS data of tryptic peptides from bovine serum albumin; The tryptic peptides are separated within a 60-min gradient. The mass spectrum of HPLC fraction eluting at 16–17 min (labeled with a red bar in A) shows two peaks (in red) with different charge states corresponding to the tryptic peptides with two interchain disulfide linkages. Inset: The zoom-in spectra of detected disulfide-linked peptide peaks, used to identify the monoisotopic peaks and charge states.

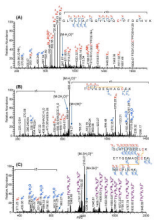


Figure 4.

Representative ESI MS/MS data for the disulfide-linked tryptic peptides from bovine serum albumin; (A) peptide with a free cysteine residue; (B) peptide with an internally linked disulfide bond. Fragment ions, y and b ions resulting from the cleavage of the intradisulfide bond are designated $y_{n\alpha}$ and $b_{n\alpha}$; (C) peptides with two interchain disulfide linkages."

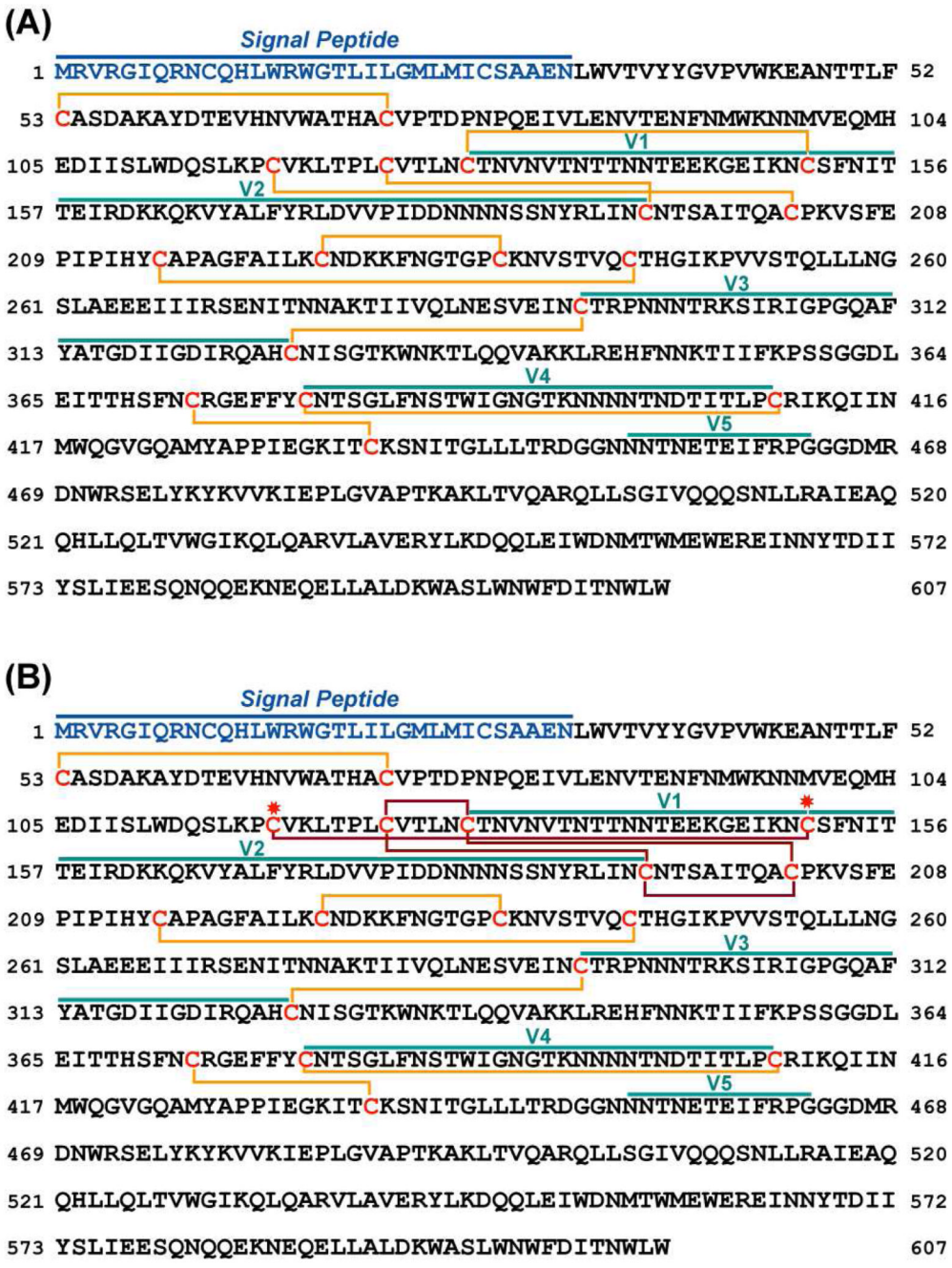


Figure 5.

Overview of (A) Expected disulfide-linkages based on the disulfide bond arrangement of a previous analysis of a monomeric form of HIV-1 envelope protein, gp120 expressed in CHO cells[2] and (B) experimentally determined disulfide linkages in CON-S gp140 ΔCFI. Disulfide linkages are indicated with a solid line and cysteine residues are shown in red and the Env signal peptide sequence is shown in blue. The cysteine residues labeled with a star in (B) represent free cysteines. The variable regions (V1 to V5) are labeled.

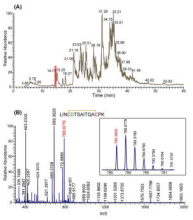


Figure 6.
Representative (A) reversed phase HPLC chromatogram and (B) LC/ESI-FTICR MS data of tryptic peptides from CON-S gp140 ΔCFI; The tryptic peptides are separated within a 60-min gradient. The MS spectrum of HPLC fraction eluting at 14–15 min (labeled with a red bar in A) shows a peak at *m/z* 788.8779 (in red) corresponding to the disulfide-linked tryptic peptides. Inset: The zoom-in spectrum of the detected disulfide-linked peptide peak, which is used to identify the monoisotopic peak and charge state.

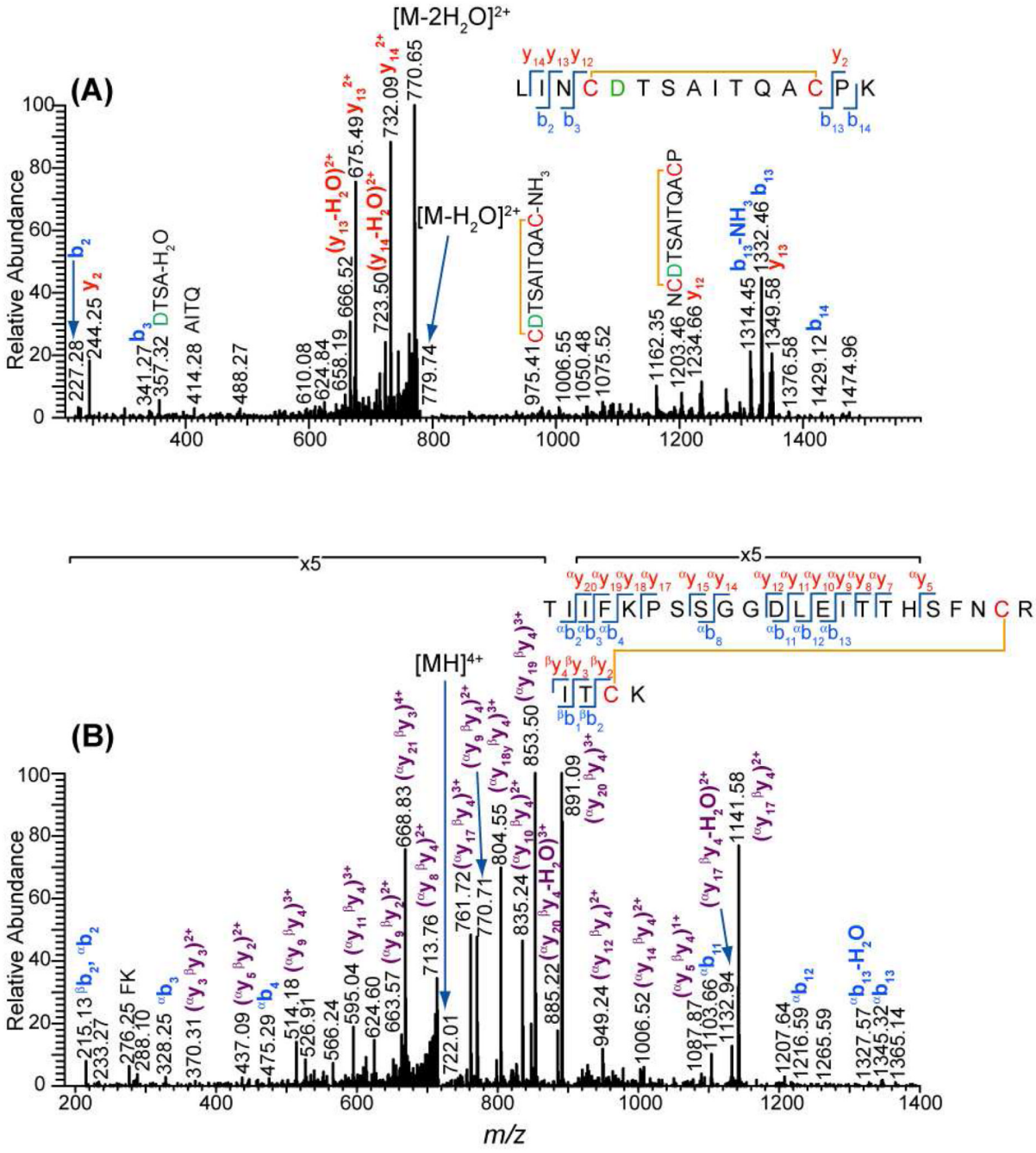


Figure 7.

Representative ESI-MS/MS data for the disulfide-linked tryptic peptides from CON-S gp140 ΔCFI (A) peptide with an internally linked disulfide bond; (B) disulfide-linked peptides within the V4 loop.

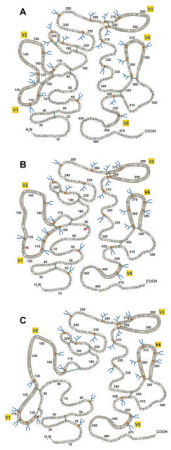


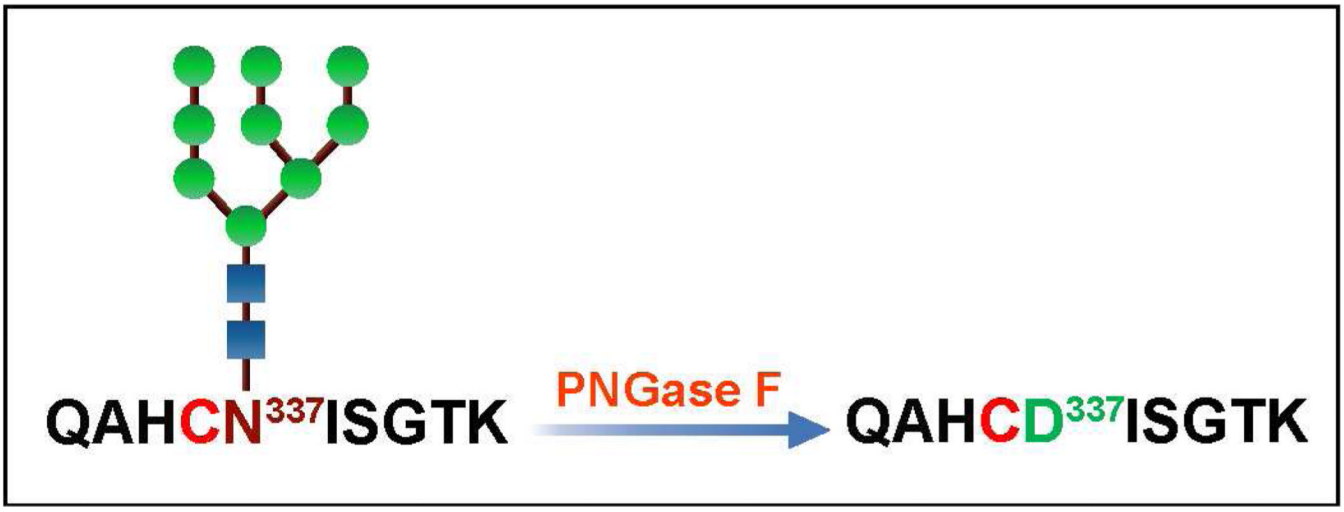
Figure 8.

Pictorial representations of the disulfide connectivity in (A) a previously analyzed Env protein[2] and (B) and (C) the Env analyzed in this study. The Env in this study showed heterogeneity in its disulfide bonding pattern in the V1 and V2 regions, so two representations, with the two different patterns, are shown. The pictorial representation was adapted from the known schematic representation of the disulfide bonding pattern of a III_B HIV-1 isolate gp120 expressed in CHO cells by Leonard, CK et. al.[2] Each sphere represents the amino acid residues of CON-S gp140 ΔCFI arranged according to the amino acid sequence. The disulfide linkages as well as the glycans attached to potential glycosylation sites are shown. Complex glycans are represented by

while high mannose/hybrid glycans are represented by

. S-pyridylethyl group on Cys residues is indicated by

.



Scheme 1.
The enzymatic release of glycans using PNGase F converts asparagine (N) in the consensus sequence, NXT/S, to aspartic acid (D), resulting in a mass shift of 1 Da.

Table 1

Summary of the disulfide linkages in bovine serum albumin

Cys-containing peptides from BSA	Theoretical *m/z*	Experimental *m/z*	Mass Error (ppm)	Charge State
GLVLIAFSQYLQQC*PFDEHVK	847.4384	847.4376	0.9	3+
	635.8306	635.8303	0.5	4+
TCVADESHAGCEK	1347.5303	1347.5283	1.5	1+
	674.2688	674.2680	1.2	2+
SLHTLFGDELCK ETYGDMADCCEK NECFLSHK	1233.5240	1233.5219	1.7	3+
	925.3948	925.3934	1.5	4+
LKPDPNTLCDEFK	1283.5928	1283.6040	8.7	3+
YNGVFQECCQAEDK	962.9464	962.9460	0.4	4+
GACLLPK	770.5586	770.5576	1.3	5+
CASIQK ECCHGDLLECADDR	1111.9503	1111.9495	0.7	2+
	741.6359	741.6351	1.1	3+
	556.4788	556.4783	0.9	4+
YICDNQDTISSK ECCDKPLLEK SHCIAEVEK	1191.8787	1191.8763	2.0	3+
	894.1609	894.1600	1.0	4+
	715.5301	715.5289	1.7	5+
	596.4430	596.4403	4.5	6+
DVCK EYEATLEECCAK DDPHACYSTVFDK	1115.4628	1115.4625	0.2	3+
	836.8489	836.8484	0.6	4+
QNCDQFEK	1242.5506	1242.5502	0.3	3+
CCTKPESER	932.1648	932.1635	1.4	4+
MPCTEDYLSLILNR	745.9333	745.9327	0.8	5+
	621.7789	621.7783	1.0	6+

Cys-containing peptides from BSA	Theoretical *m/z*	Experimental *m/z*	Mass Error (ppm)	Charge State
LCVLHEK CCTESLVNR RPCFSALTPDETYVPK	1228.5944 921.6977 737.5596 614.8009	1228.5937 921.6966 737.5589 614.8007	0.6 1.2 0.9 0.3	3+ 4+ 5+ 6+
LFTFHADICTLPDTEK CCAADDK EACFAVEGPK	905.9068 724.9269	905.9069 724.9261	0.1 1.1	4+ 5+

* Free cysteine is alkylated with 4-vinylpyridine

Table 2

Glycosylation site occupancy on the Cys-containing tryptic peptides of CON-S gp140 ΔCFI

CON-S gp140 Δ CFI	Number of potential glycosylation sites	Number of sites occupied	Theoretical deglycosylated peptide mas (*m/z*)
EAN^{48}TTLFCASDAK	1	1	1371.6096
AYDTEVHNVWATHACVPTDPNPQEIVLEN^{87}VTENFNMWK	1	1	4413.0226
CNDKKFN^{237}GTGPCK	1	1	1412.6297
QAHCN^{337}ISGTK	1	1	1059.4888
GEFFYCN^{391}TSGLFN^{397}STWIGN^{403}GTK	3	3	2446.0543
LTPLCVTLN^{129}CTNVN^{135}VTN^{138}TTN^{141}NTEEK	4	2 and 3	2738.2859 and 2739.2699
N^{155}CSFN^{159}ITTEIR	2	1 and 2	1298.6045 and 1299.5885
LINCN^{201}TSAITQACPK	1	0 and 1	1576.7822 and 1577.7662
N^{245}VSTVQCTHGIKPVVSTQLLLN^{266}GSLAEEEIIIR	2	1 and 2	3562.9148 and 3563.8989
TIIVQLN^{293}ESVEIN^{299}CTRPNN^{305}NTR	3	1 and 2	2529.2726 and 2530.2566
NNN^{413}NTN^{416}DTITLPCR	2	1 and 2	1590.7177 and 1591.7017

Table 3

Summary of the disulfide linkages in CON-S gp140 ΔCFI detected by LC/ESI-FTICR MS

Disulfide-linked Peptides in the C1, C2, C3, and C4 regions and V3 and V4 loops	Theoretical *m/z*	Experimental *m/z*	Mass Error (ppm)	Charge State
EADTTLFCASDAK AYDTEVHNVWATHACVPTDPNPQEIVLEDVTENFNMWK	1446.1578 1157.1277	1446.1536 1157.1269	2.9 0.7	4+ 5+
NNMVEQMHEDIISLWDQSLKPCVK NCSFDITTEIR	1384.9881 1038.9929	1384.9917 1039.0003	2.6 7.1	3+ 4+
NNMVEQMHEDIISLWDQSLKPCVK DCSFDITTEIR	1384.3161 1039.2389	1385.3240 1039.2472	5.7 8.0	3+ 4+
LTPLCVTLDCTNVDVTNTTNNTEEK LINCDTSAITQACPK	1437.6760 1078.5088	1437.6745 1078.5032	1.0 5.2	3+ 4+
VSFEPIPIHYCAPAGFAILK DVSTVQCTHGIKPVVSTQLLLDGSLAEEEIIIR	1434.2623 1147.6113	1434.2772 1147.6259	10.4 12.7	4+ 5+
TIIVQLDESVEINCTRPNDNTR QAHCDISGTK	1196.2457 897.4361	1196.2547 897.4437	7.5 8.5	3+ 4+
TIIVQLDESVEIDCTRPNDNTR QAHCDISGTK	1196.5737 897.6821	1196.5745 897.6865	0.7 4.9	3+ 4+
THFKPSSGGDLEITTHSFNCR ITCK	962.1493 721.8638	962.1529 721.8658	3.7 2.8	3+ 4+
NNNNTDDTITLPCR GEFFYCDTSGLFDSTWIGDGTK	1345.2545 1009.1927	1345.2673 1009.2017	9.6 8.9	3+ 4+
NNDNTDDTITLPCR GEFFYCDTSGLFDSTWIGDGTK	1345.5825 1009.4387	1345.5793 1009.4349	2.4 3.8	3+ 4+
Peptides that are capped with vinylpyridine	**Theoretical *m/z***	**Experimental *m/z***	**Mass Error (ppm)**	**Charge State**
NC*SFDITTEIR	702.3348	702.3346	0.3	2+
NNMVEQMHEDIISLWDQSLKPC*VK	988.1468	988.1543	7.5	3+

Disulfide-linked Peptides in the C1, C2, C3, and C4 regions and V3 and V4 loops	Theoretical *m/z*	Experimental *m/z*	Mass Error (ppm)	Charge State
Peptides that are internally linked	**Theoretical *m/z***	**Experimental *m/z***	**Mass Error (ppm)**	**Charge State**
LTPLCVTLDCTNVDVTNTTNNTEEK	1368.6388	1368.6842	6.9	2+
	912.7616	912.7675	6.5	3+
	787.8869	787.8895	3.3	2+
	1575.7505	1575.7555	3.2	1+
	788.3789	788.3804	1.9	2+
(Pyro-cmC)NDKKFDGTGPCK	697.2974	697.3009	5.1	2+

* Glycosylated asparagines in NXT/S are converted to aspartic acid upon deglycosylation and labeled in green. Nonglycosylated Asn's are labeled in dark red.