DOI 10.1002/pmic.201200040 *Proteomics* 2012, *12*, 2746–2752

RESEARCH ARTICLE

# Analysis of secondary structure in proteins by chemical cross-linking coupled to MS

*Mariana Fioramonte[1], Aline Mara dos Santos[2], Sean McIlwain[3], William S. Noble[3], Kleber G. Franchini[2] and Fabio C. Gozzo[1]*

[1] Chemistry Institute University of Campinas, Campinas, SP, Brazil
[2] Brazilian National Laboratory for Biosciences, Campinas, SP, Brazil
[3] Department of Genome Sciences University of Washington, Seattle, WA, USA

Received: February 2, 2012
Revised: June 13, 2012
Accepted: June 14, 2012

Chemical cross-linking is an attractive technique for the study of the structure of protein complexes due to its low sample consumption and short analysis time. Furthermore, distance constraints obtained from the identification of cross-linked peptides by MS can be used to construct and validate protein models. If a sufficient number of distance constraints are obtained, then determining the secondary structure of a protein can allow inference of the protein's fold. In this work, we show how the distance constraints obtained from cross-linking experiments can identify secondary structures within the protein sequence. Molecular modeling of alpha helices and beta sheets reveals that each secondary structure presents different cross-linking possibilities due to the topological distances between reactive residues. Cross-linking experiments performed with amine reactive cross-linkers with model alpha helix containing proteins corroborated the molecular modeling predictions. The cross-linking patterns established here can be extended to other cross-linkers with known lengths for the determination of secondary structures in proteins.

**Keywords:**
Alpha helix / Beta sheet / Cross-linking / MS / Protein secondary structure / Technology

## 1 Introduction

Chemical cross-linking is a very attractive technique for the study of protein structure, especially in the absence of data from high-resolution methods such as nuclear magnetic resonance (NMR) and protein crystallography. The cross-linking technique is based on the formation of covalent bonds between amino acid residues that are near to one another in the protein structure. Cross-linked peptides can then be identified by shotgun proteomics analysis, which gives information for residues within the protein structure that should be closer in space than the cross-linker reagent's spacing arm. By using cross-linking, several different structural properties can be determined:

(1) modification of residues indicates that they are solvent exposed;
(2) interprotein cross-linked peptides (composed of peptides from two different proteins) reveal the topology of a multichain complex, and the distance constraints of these interprotein cross-linked peptides can be used to generate a model of the complex; and
(3) intraprotein cross-links (composed of peptides from the same chain) reveal information about the folding of an individual protein.

Due to its attractive features, including low sample consumption, short analysis time, and relative ease of use, cross-linking has been used to study multiple proteins and protein complexes. By far, the most common uses of cross-linking are either to determine the topology or to identify the interaction region between two proteins in a complex. Several examples have been shown in the literature for these applications [1–5].

An alternative potential application of cross-linking/MS is to use the distance constraints provided by intraprotein

**Correspondence:** Professor Fabio Cesar Gozzo, Chemistry Institute, University of Campinas, CP6154, Campinas, SP 13083-970, Brazil
**E-mail:** fabio@iqm.unicamp.br
**Fax:** +55-19-35213008

**Abbreviation: CD**, circular dichroism

**Colour Online:** See the article online to view Figs. 1–4 in colour.

cross-links to reveal structural information within a single protein. Assuming that the general type of fold is known and that enough cross-links are found, the corresponding set of distance constraints can be used to derive the precise fold of the protein. This approach is similar to NMR protein structure determination methods, where a set of distance constraints are obtained by NOESY (nuclear Overhauser effect spectroscopy) experiments for use in molecular modeling. Currently, the difficulty in the case of cross-linking is the relatively small number of cross-links that can be obtained for a specific protein. Young et el. [6] have suggested that if the number of cross-links is higher than 10% of the number of residues, then a fold may be determined. However, other theoretical studies suggest that this number may be higher for some specific folds [7]. As far as we are aware, the only work to use cross-linking to model a protein structure has been the one by Young et al.

Another potential use of cross-linking is the analysis of secondary structure. Currently, the most common methods for secondary structure determination are circular dichroism [8] (CD) and homology modeling [9,10]. Although CD is very suitable for determining the overall presence of alpha helices and beta sheets, the data from these experiments cannot localize these individual structures in the protein sequence. Several secondary structure prediction programs are currently available [11–13], but their accuracy is somewhat limited and care should be taken in interpreting the results.

Because alpha helices and beta sheets force the amino acid side chains to have a specific orientation, the distances between side chains are restricted to a relatively short range. Thus, cross-linking can be used as a ruler, and the formation of a cross-link can be indicative of a specific secondary structure. Certain residues have a tendency to form a specific secondary structure: amino acids with bulky side chains (tryptophan, tyrosine, phenylalanine, isoleucine, valine, and threonine) prefer to adopt beta strand conformation [14], whereas methionine, alanine, leucine, glutamic acid, and lysine prefer to form alpha helices [15].

Currently, lysine-specific cross-linkers (based on *N*-hydroxy-succinimide esters) are the most commonly used cross-linkers, making them very suitable for alpha helix probing. If cross-linkers with other specificities are used, such as photoactivable cross-linkers based on diazirines [16] or benzophenones [17], then possible beta sheets can also be interrogated. In this work, we explore the use of cross-linking to probe the presence of alpha helices and beta sheets by molecular modeling and lysine-specific cross-linking experiments for alpha helices.

## 2 Materials and methods

### 2.1 Materials

The proteins hemoglobin from bovine blood, myoglobin from horse heart, and ubiquitin from bovine red blood cells were obtained from Sigma-Aldrich (St Louis, MO, USA). GST-Myosin (~48 KDa) encoding the C-terminal region of the α-myosin heavy chain cloned in pGEX5 × 2 (GE Healthcare, Piscataway, NJ, USA) was also used [18]. This construct was expressed in *Escherichia coli* Bl21 (DE3) and purified by affinity chromatography (HisTrap HP 5 mL, GE Healthcare) and by exclusion size chromatography (HiLoad 26/60 Superdex 200 pg, GE Healthcare). Dithiothreitol and iodoacetamide were obtained from Pierce. The cross-linker disuccinimidyl suberate (DSS) was obtained from Sigma-Aldrich (St Louis, MO, USA). Sequencing grade modified trypsin was obtained from Promega, Madison, WI, USA.

### 2.2 Chemical cross-linking reaction

Chemical cross-linking was performed as previously described [16]. The proteins were solubilized in 50 mM pH 8.0 phosphate buffer to give a final volume of 500 μL and concentration of 1 mg/mL. DSS was dissolved in *N*,*N*-dimethylformamide (10 mg/mL) and immediately added to the protein solutions in a 50:1 molar excess. Reactions were allowed to stand at room temperature for 2 h. When alkylation of thiol groups was necessary, it was performed by reduction with 250 mM dithiothreitol solution at a final concentration of 60 mM, for 30 min at room temperature, followed by a reaction with 250 mM iodoacetamide at a final concentration of 60 mM, kept in the dark at room temperature for 30 min. Trypsin (sequencing grade modified porcine pancreas) was added to a final concentration of 1:20 (w/w), and the digestion was carried out for 16 h at 37°C.

### 2.3 MS analysis

LC-ESI-QTOF-MS/MS experiments on the tryptic digests of the cross-linked proteins were performed on a Waters nano-Acquity UPLC (ultra performance liquid chromatography) coupled to a Waters Synapt HDMS mass spectrometer. The UPLC system was fitted with a Waters Symmetry C18 trap column (20 mm × 180 μm id; 5 μm particle size), followed by a Waters BEH130 C18 analytical column (100 mm × 100 μm id; 1.7 μm particle size). Samples were injected and washed on the trapping column for 3 min with 97:3 $H_2O$/MeCN with 0.1% formic acid, at a 5 μL/min flow rate, and then eluted with a gradient of 97:3–30:70% $H_2O$/MeCN with 0.1% formic acid, at a flow rate of 1 μL/min. Analysis was performed using data-dependent acquisition.

### 2.4 Identification of cross-linking products from LC-MS/MS experiments

The raw data files from the LC-MS/MS runs were processed using Mascot Distiller (Matrix Science Ltd.). The inter- and intraprotein cross-links were identified using the program Crux search-for-xlinks [19] followed by manual validation. The search parameters were as follows: precursor tolerance 0.1

 

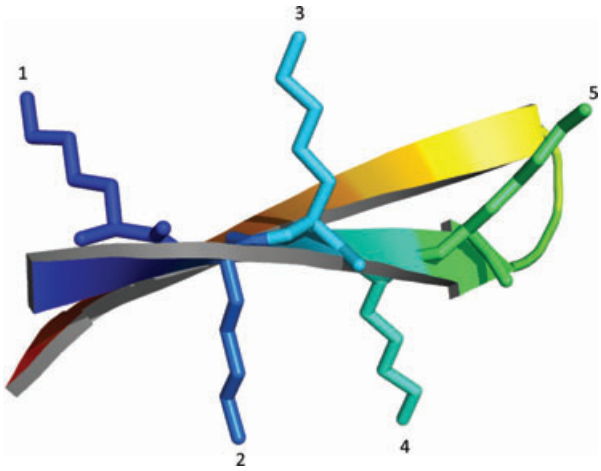


**Figure 1.** Structure of a polypeptide in twisted beta sheet conformation showing lysine residues at positions 1, 3, and 5.

Da, missed cleavages 2, variable modification 156.07 Da on lysines (corresponding to dead ends), up to two modifications per peptide.

### 2.5 Molecular dynamics

Molecular dynamics simulations were performed using the software NAMD 2.8. For alpha helix, eleven hypothetical poly-alanine alpha helix (20 residues in length) containing two lysine residues were modeled. In these peptides, one lysine was always placed as the second residue, whereas the other lysine was placed varying its position from residue 3 till position 13, resulting in positions 2 till 12 relative to the first lysine. For beta strands, a two strand beta sheet composed of MKIKVKTKTGKTITLEV sequence was modeled. NVT (constant temperature and volume) ensemble was used with a water box with enough chloride ions to neutralize the system, using the topology file par_all27_prot_lipid, during 4 ns.

### 2.6 Topological distances

The Xwalk [21] software was used to calculate the topological distances from the modeled alpha helices and the hypothetical polylysine beta sheet. Topological distances were calculated between the nitrogen atoms in the lysine side chains.

## 3 Results and discussion

### 3.1 Structure of beta sheets and alpha helices

Beta strands have a linear structure, where the backbone is arranged in an almost straight line, and the side chains are located alternately on each side of the strand. The strand is usually twisted along the main backbone line; and along with the other strands, they form a beta sheet, where the side chains are pointed upwards or downwards. A hypothetical polylysine in beta strand conformation is shown in Fig. 1. As shown, the cross-linkable residues will be the ones on the same side of the beta strand, that is, among the odd or even numbered residues, because residues on different side of the strand should not be cross-linkable using a short- or medium-sized cross-linker (spacing arm <15 Å).

Alpha helices have 3.6 residues per turn, which translates into an angle of 100 degrees between residues, with the side chains of each residue pointing outwards (Fig. 2). This configuration produces geometric features that can be used to derive several cross-linking rules from the sequence positions. For a perfect alpha helix (Fig. 2), some side chains will be on the same side of the helix, facilitating the cross-linking, whereas others will lie on opposite sides of the helix, making cross-linking unlikely for a short- or medium-range cross-linker (<15 Å). Interestingly, in alpha helices, some residues that are far apart in the primary sequence may still have an appropriate geometry for cross-linking, i.e., they may be close in space (parallel). This is the case for residues 8 and 1 in the alpha helix, which could be cross-linked to residue 1 if a

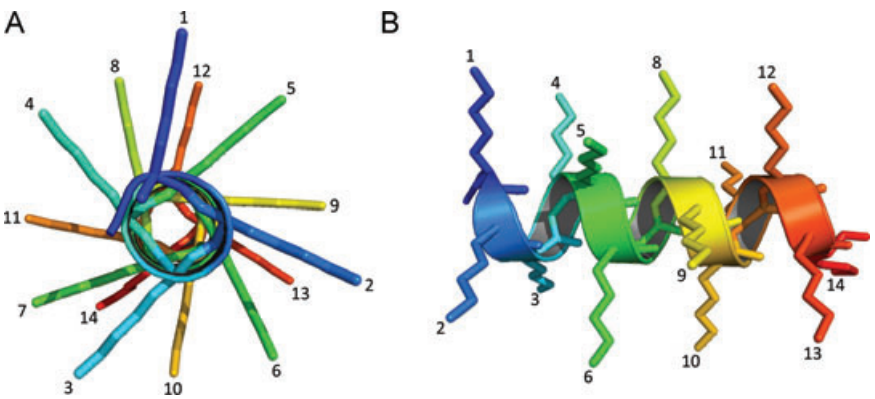


**Figure 2.** (A) A hypothetical lysine 14-mer in the alpha helix conformation. (B) Side view of the same alpha helix.

 

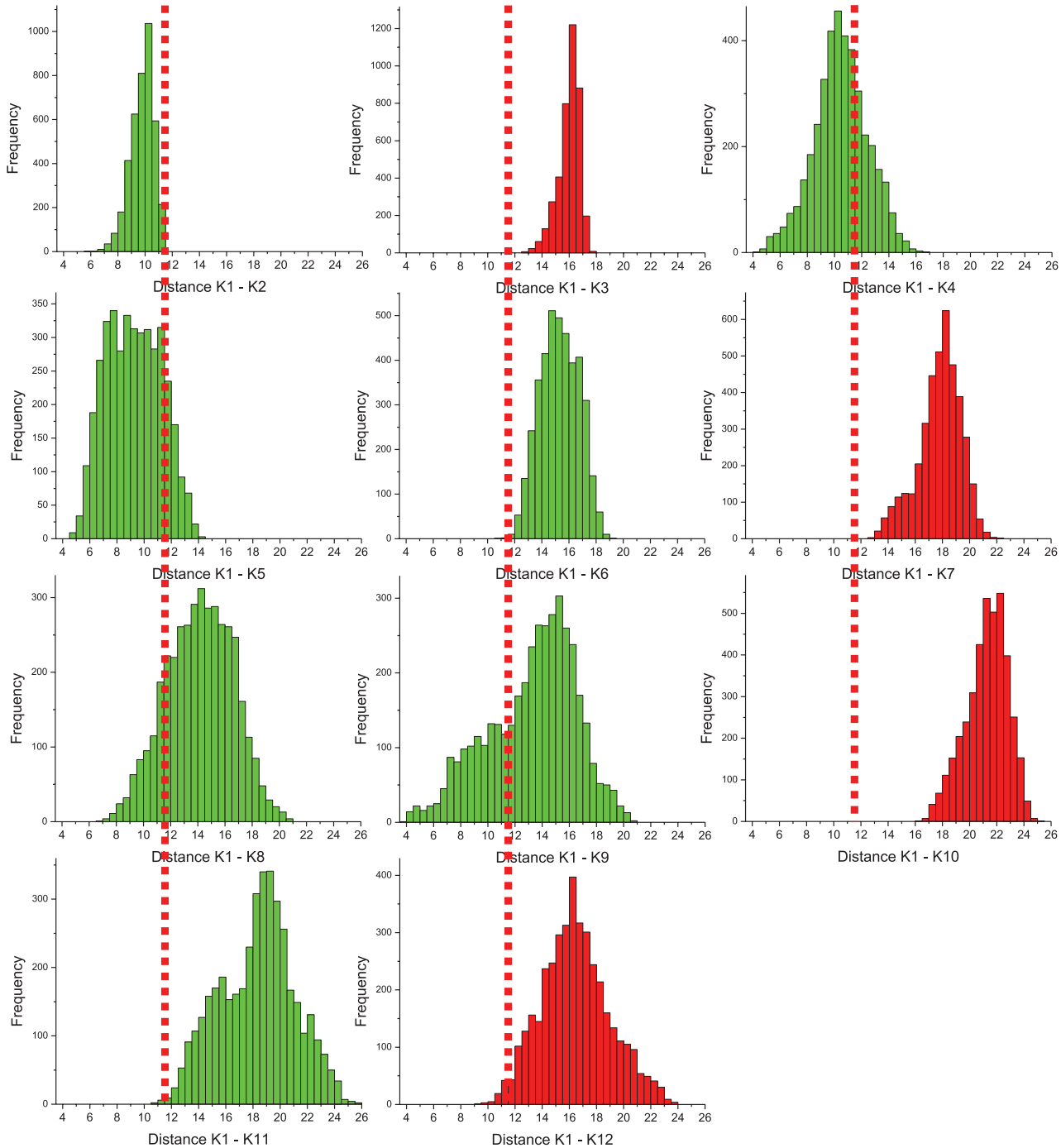


**Figure 3.** Histograms of direct (Euclidean) distances between amino groups of lysine side chains for alpha helices containing lysine residues at different relative positions. The dashed line corresponds to arm length of DSS/BS3 cross-linkers (11.5 Å).

cross-linker with arm length of 12 Å was used (Fig. 2). In a beta strand, on the other hand, a large distance in the primary sequence would make the amino acids much farther apart in space (Fig. 1).

### 3.2 Molecular dynamics simulations

For a comprehensive analysis of cross-linking possibilities, three factors should be taken into account:

 

**Table 1.** Topological distances between the amino group of lysine residues at the conformation with the shortest direct (Euclidean) distance for alpha helices and beta strands with the second lysine residue at different relative positions

| Lysine relative positions (alpha helix) | Topological distances (Å) | Lysine relative positions (beta strand) | Topological distances (Å) |
|---|---|---|---|
| K1–K2 | 10.6 | K1–K2 | 17.5 |
| K1–K3 | 19.5 | K1–K3 | 8.1 |
| K1–K4 | 9.3 | K1–K4 | 19.5 |
| K1–K5 | 5.5 | K1–K5 | 17.6 |
| K1–K6 | 10.9 | K1–K6 | 19.6 |
| K1–K7 | 20.6 | K1–K7 | 21.7 |
| K1–K8 | 7.3 | K1–K8 | 25.9 |
| K1–K9 | 7.6 | K1–K9 | 36.2 |
| K1–K10 | 22.1 | – | – |
| K1–K11 | 11.5 | – | – |
| K1–K12 | 12.9 | – | – |

(1) The cross-linker spacing length will dictate the maximum distance between the cross-linkable residues. For simplicity, all analyses here will be based on the most commonly used cross-linkers, DSS, and BS3. Both cross-linkers have a maximum spacer length of 11.5 Å.
(2) The dynamics (flexibility) of both backbone and side chain has to be considered, especially in the case of lysine residues because they are long chains and can, in principle, acquire very different conformations.
(3) The distance between two residues should be smaller than the cross-linker spacing length. This distance is not a direct (Euclidean) distance, but topological, since the cross-linker should be able to link the residues across the protein surface.

To carry out a comprehensive analysis including these three factors, molecular dynamics were performed in a model alpha helix composed mainly of alanines with two lysines in sequential positions (Fig. 3 and Table 1) as well as a beta sheet with lysines at positions 1, 3, and 5. The most common conformations of these structures were then analyzed by the Xwalk software to obtain topological distances.

For an alpha helix, residues located on opposite sides of the helix should have a longer topological distance than the ones on the same side. For example, the topological distance between lysine amino groups for K1 and K2 is approximately 17 Å, whereas the topological distance between K1 and K3 is approximately 23 Å, because the cross-linker needs to make a turn around the helix to reach K3. For K1 and K4, the distance is smaller again (around 14 Å) because the third residue nearly completes one turn, approaching the first lysine. The same pattern is observed for the following turns until residue K11, which is the farthest residue that is linkable in the protein sequence at a topological distance of 11.5 Å. From these observations, a set of selection rules can be devised for alpha helix cross-linking possibilities, considering both the side chain flexibility and the topological distances (Table 2).

**Table 2.** Cross-link selection rules for alpha helices and beta strand for cross-linkers DSS/BS3

| Lysine relative positions | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Alpha Helix | A | F | A | A | A | F | A | A | F | A | F |
| Beta Strand | F | A | F | F | F | F | F | F | F | F | F |

A, allowed; F, forbidden.

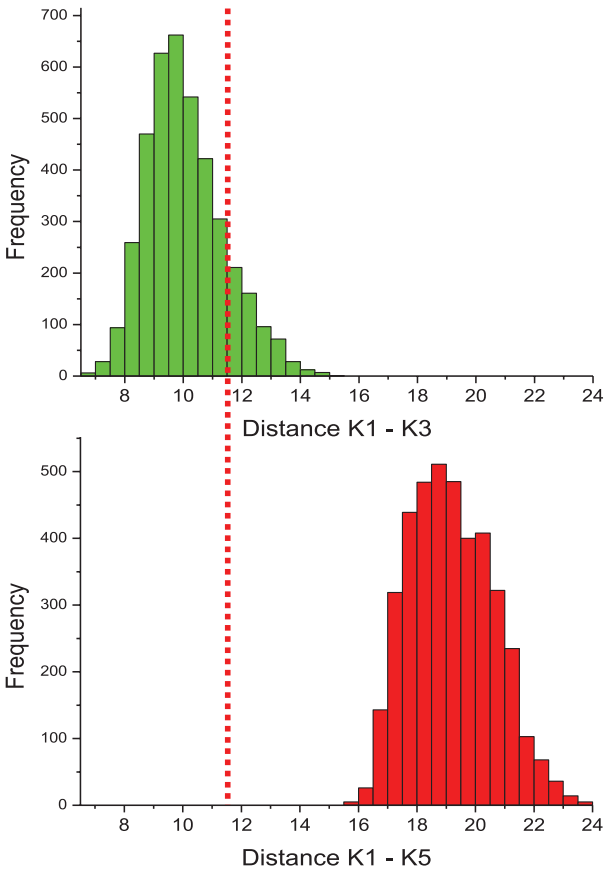


**Figure 4.** Histograms of direct (Euclidean) distances between amino groups of lysine side chains for beta strands containing lysine residues at relative positions 3 and 5. The dashed line corresponds to arm length of DSS/BS3 cross-linkers (11.5 Å).

The equivalent analysis can be performed for beta strands (Table 1, Fig. 4). Here the cross-linking possibilities are restricted to residue 3 only, because residue 2 is on the opposite side of the strand, with a topological distance of 17.5 Å, and residue 4 and all subsequent residues have a topological distance of greater than 17 Å. Hence, for a beta strand, the only available cross-link is between K1 and K3. This situation is exactly the opposite of that for the alpha helix, where cross-linking between residues 1 and 3 is forbidden due to the topological distance (>20 Å, Table 1).

 

**Table 3.** Identified interprotein cross-links with DSS

| Protein | K1 position | K2 position | K2 relative position | Euclidean distance (Å) | Topological distance (Å) |
|---|---|---|---|---|---|
| Ubiquitin | 229 | 233 | 5 | 6.4 | 7.1 |
| Myoglobin | 61 | 62 | 2 | 5.3 | 6.2 |
| Myoglobin | 145 | 147 | 3 | 7.5 | 9.3 |
| α-hemoglobin | 11 | 16 | 6 | 9.8 | 15.8 |
| Myosin | 42 | 45 | 4 | 5.7 | 8.0 |
| Myosin | 45 | 46 | 2 | 5.4 | 5.6 |
| Myosin | 45 | 50 | 6 | 9.8 | 11.5 |
| Myosin | 67 | 74 | 8 | 10.4 | 12.3 |
| Myosin | 94 | 95 | 2 | 5.3 | 5.5 |
| Myosin | 155 | 159 | 5 | 7.9 | 10.4 |
| Myosin | 396 | 384 | 11 | – | – |

Myosin structure was obtained from ref. [1]. No structure is available for the sequence involved in myosin cross-link 396–384.

### 3.3 Cross-linking experiments with DSS

To confirm the cross-linking possibilities predicted by the molecular modeling, cross-linking experiments with DSS and the alpha helix rich proteins ubiquitin, myoglobin, hemoglobin, and myosin were performed, looking for intraprotein cross-links. Several of these cross-links were identified by the Crux search-for-xlinks software, followed by manual validation (Table 3 and Supporting Information Figs. S1 and S2).

The data shown in Table 3 fit perfectly with the predicted cross-linking possibilities for alpha helices (Table 2). The predicted cross-links at positions 2, 4, 5, 6, 8, and 11 were found in the experiments. The only exception is the cross-link between residues 145 and 147 from myoglobin. This cross-link (Fig. 5) is well characterized by MS and it seems to violate the predicted distances for cross-links at relative position 3. A closer inspection of the structure of myoglobin reveals, however, that these lysine residues belong to the end of alpha helix 8, with lysine 147 being the last residue of the helix. Moreover, this residue is the third from last residue in the protein, making this part of the protein flexible. Indeed, the analysis of the NMR structure (PDB ID 1MYF) of this part of the protein shows a reasonable flexibility of the protein C-terminus, distorting it and allowing the two lysines to have a shorter topological distance (12 Å). Therefore, it is quite likely that this part of the protein is flexible enough to acquire conformations with a topological distance between lysine residues 145 and 147 that would be within the reach of the DSS cross-linker. Thus, the experimental cross-linking data corroborates the predicted selection rules derived from molecular modeling.

The case of the cross-link between lysines 145 and 147 from myoglobin shows that the selection rules presented here can be violated by chain flexibility or secondary structure distortion, but molecular modeling should be able to account for these artifacts. Another possible caveat is the cross-linking in random coil regions. Because random coils are usually very

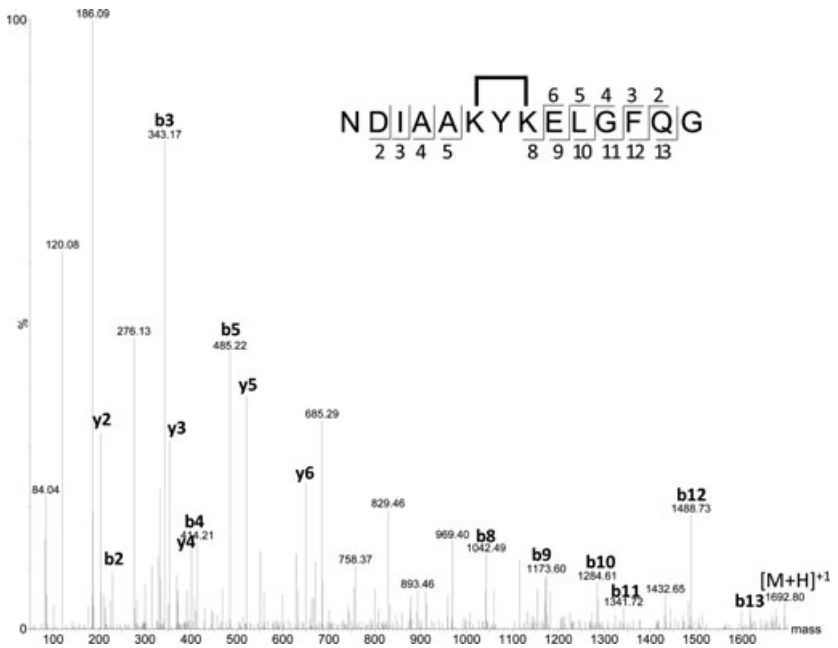


**Figure 5.** MS/MS spectrum of intramolecular cross-link between lysines 145 and 147 from myoglobin.

 

flexible and do not enforce any specific orientation of residues side chains, practically any cross-linking may be allowed in these secondary structures. This phenomenon should always be taken into account when creating a protein model.

## 4 Conclusions

We showed that chemical cross-linking coupled to MS can be used to generate structural information that correlates with protein's secondary structure. The presence of alpha helices and beta strands can be determined using the selection rules found for each secondary structure. For DSS/BS3 cross-linkers, alpha helices allow cross-links between lysines at relative positions 2, 4, 5, 6, 8, 9, and 11, whereas beta strands only allow cross-links between residues at a relative position 3. These complementary selection rules are useful for distinguishing alpha helices from beta strands.

The selection rules presented here can be extended to other cross-linkers with known spacing lengths. Furthermore, although beta sheets were not experimentally probed here due to the lack of lysine residues, the use of photoactivable, non-specific cross-linkers can be appropriate for determining the secondary structure. Thus, the information obtained by cross-linking holds great promise for determining secondary structures within a protein sequence through the use of all of the analytical advantages of mass spectrometric detection.

*The authors would like to thank Paulo de Souza for useful discussions on the molecular dynamics calculations and the financial support from FAPESP, CNPq, Instituto Nacional de Ciencia e Tecnologia em Bioanalitica (INCTBio) and Genoprot Program for Intracelular Peptides.*

*The authors have declared no conflict of interest.*

## 5 References

[1] Santos, A. M., Schechtman, D., Cardoso, A. C., Clemente, C. F., Silva, J. C., Fioramonte, M., Pereira, M. B. M. et al., FERM domain interaction with myosin negatively regulates FAK in cardiomyocyte hypertrophy. *Nat. Chem. Biol.* 2012, *8*, 102–110.

[2] Dimova, K., Kalkhof, S., Pottratz, I., Ihling, C. et al., Structural insights into the Calmodulin-Munc13 interaction obtained by cross-linking and mass spectrometry. *Biochemistry* 2009, *48*, 5908–5921.

[3] Trnka, M. J., Burlingame, A. L., Topographic dtudies of the GroEL-GroES chaperonin complex by chemicalcross-linking using diformyl ethynylbenzene. *Mol. Cell. Proteomics* 2010, *9*, 2306–2317.

[4] Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C. et al., Architecture of the RNA polymerase II–TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO J.* 2010, *29*, 717–726.

[5] Zhang, H., Tang, X., Munske, G. R., Tolic, N. et al., Identification of protein-protein interactions and topologies in living cells with chemical cross-linking and mass spectrometry. *Mol. Cell. Proteomics* 2009, *8*, 409–420.

[6] Young, M. M., Tang, N., Hempel, J. C., Oshiro, C. M. et al., High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc. Natl. Acad. Sci. USA* 2000, *97*, 5802–5806.

[7] Chen, Y., Ding, F., Dokholyan, N. V., Fidelity of the protein structure reconstruction from inter-residue proximity constraints. *J. Phys. Chem. B* 2007, *111*, 7432–7438.

[8] Wallace, B. A., Protein characterisation by synchrotron radiation circular dichroism spectroscopy. *Q. Rev. Biophys.* 2009, *42*, 317–370.

[9] Liu, H. L., Hsua, J. P., Recent developments in structural proteomics for protein structure determination. *Proteomics* 2005, *5*, 2056–2068.

[10] Chou, K. C., Some remarks on protein attribute prediction and pseudo amino acid composition. *J. Theor. Biol.* 2011, *273*, 236–247.

[11] Jones, D. T., Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 1999, *292*, 195–202.

[12] Pollastri, G., McLysaght, A., Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 2005, *21*, 1719–1720.

[13] Adamczak, R., Porollo, A., Meller, J., Combining prediction of secondary structure and solvent accessibility in proteins. *Proteins* 2005, *59*, 467–475.

[14] Kim, C. A., Berg, J. M., Thermodynamic β-sheet propensities measured using a zinc-finger host peptide. *Nature* 1993, *362*, 267–270.

[15] Pace, C. N., Scholtz, J. M., A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 1998, *75*, 422–427.

[16] Gomes, A. F., Gozzo, F. C., Chemical cross-linking with a diazirine photoactivatable cross-linker investigated by MALDI- and ESI-MS/MS. *J. Mass Spectrom.* 2010, *45*, 892–899.

[17] Krauth, F., Ihling, C. H., Rüttinger, H. H., Sinz, A., Heterobifunctional isotope-labeled amine-reactive photo-cross-linker for structural investigation of proteins by matrix-assisted laser desorption/ionization tandem time-of-flight and electrospray ionization LTQ-Orbitrap mass spectrometry. *Rapid Commun. Mass Spectrom.* 2009, *23*, 2811–2818.

[18] Fonseca, P. M., Inoue, R. Y., Kobarg, C. B., Crosara-Alberto, D. P. et al., Targeting to C-terminal myosin heavy chain may explain mechanotransduction involving focal adhesion kinase in cardiac myocytes. *Circ. Res.* 2005, *96*, 73–81.

[19] McIlwain, S., Draghicescu, P., Singh, P., Goodlett, D. R. et al., Detecting cross-linked peptides by searching against a database of cross-linked peptide pairs. *J. Proteome Res.* 2010, *9*, 2488–2495.

[20] Phillips, J. C., Braun, R., Wang, W., Gumbart, J. et al., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 2005, *26*, 1781–1802.

[21] Kahraman, A., Malmström, L., Aebersold, R., Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* 2011, *27*, 2163–2164.

 