

Published in final edited form as:

Anal Chim Acta. 2008 October 3; 627(1): 117–128. doi:10.1016/j.aca.2008.05.074.

Nested Arg-specific bifunctional crosslinkers for MS-based structural analysis of proteins and protein assemblies

Qingrong Zhang, Elizabeth Crosland, and Daniele Fabris *

University of Maryland Baltimore County, Baltimore, MD

Abstract

The combination of chemical probing and high-resolution mass spectrometry constitutes a powerful alternative for the structural elucidation of biomolecules possessing unfavorable size, solubility, and flexibility. We have developed nested Arg-specific bifunctional crosslinkers to obtain complementary information to typical Cys- and Lys-specific reagents available on the market. The structures of 1,4-phenyl-diglyoxal (PDG) and 4,4'-biphenyl-diglyoxal (BDG) include two identical 1,2-dicarbonyl functions capable of reacting with the guanido group of Arg residues in proteins, as well as the base-pairing face of guanine in nucleic acids. The reactive functions are separated by modular spacers consisting of one or two benzene rings, which confer greater rigidity to the crosslinker structure than it is afforded by typical aliphatic spacers. Analysis by electrospray ionization (ESI) Fourier transform ion cyclotron resonance (FTICR) mass spectrometry has shown that the probes provide both mono- and bifunctional products with model protein substrates, which are stabilized by the formation of diester derivatives in the presence of borate buffer. The identification of crosslinked sites was accomplished by employing complementary proteolytic procedures and peptide mapping by ESI-FTICR. The results showed excellent correlation with the solvent accessibility and structural context of susceptible residues, and highlighted the significance of possible dynamic effects in determining the outcome of crosslinking reactions. The application of nested reagents with different spacing has provided a new tool for experimentally recognizing flexible regions that may be involved in prominent dynamics in solution. The development of new bifunctional crosslinkers with diverse target specificity and different bridging spans is expected to facilitate the structure elucidation of progressively larger biomolecular assemblies by increasing the number and diversity of spatial constraints available for triangulating the position of crosslinked structures in the three dimensions.

Keywords

Structural probing; structure determination; bifunctional crosslinking; *bis*-(1,2-dicarbonyls); Arg-specific labels; ESI-FTICR mass spectrometry

Chemical labeling techniques and mass spectrometric analysis constitute an ideal combination for investigating the structure and dynamics of biomolecules that exceed the size accessible by nuclear magnetic resonance, or afford inadequate crystallization [1–4]. Monofunctional footprinting targeted to specific functional groups provides valuable information about

*Corresponding author: University of Maryland Baltimore County, Department of Chemistry and Biochemistry, 1000 Hilltop Circle, Baltimore, MD 21228 USA, Tel. (410) 455-3053, Fax (410) 455-2608.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

conformational changes and binding interactions that affect surface accessibility [5,6], whereas bifunctional crosslinking offers the opportunity to detect more subtle effects that do not necessarily result in accessibility variations [7–9]. In addition, bridging between structures situated within crosslinkable distance can provide valid spatial constraints for the structural elucidation of relatively large biomolecules and their assemblies. The broad applicability of soft ionization techniques, such as matrix assisted laser desorption ionization (MALDI)[10, 11] and electrospray ionization (ESI)[12,13], and the ability to perform sequence determinations [14,15] make mass spectrometry (MS) the analytical platform of choice for the characterization of products obtained from probing reactions. For this reason, approaches based on crosslinking and MS detection have been successfully applied to individual substrates of proteic [16–24] and nucleic acid nature [25,26] and to assemblies of increased complexity involving specific protein-peptide [27,28], protein-protein [29–41], and protein-nucleic acid [42–54] interactions.

We have recently demonstrated that experimental constraints obtained from bifunctional crosslinkers can be readily utilized by established computational tools to generate all-atom models of large RNA structures, such as the ribosome frameshifting pseudoknot of feline immunodeficiency virus [25] and the genome packaging signal of human immunodeficiency virus type 1 [26]. In completing molecular models based on crosslinking data, the level of confidence in the final structures tends to increase with the actual number of available constraints, which allow for greater precision in triangulating the positions of bridged groups in the three dimensions. In addition, the availability of complementary information from different reagents enables to check for inconsistencies that may reveal possible artifacts introduced by probing protocols. A subset of experimental constraints can be excluded from the modeling procedure and employed later to validate the final structure [26]. For these reasons, a wider choice of bifunctional reagents with different target specificity and crosslinking span would help diversify the type of information and would greatly increase the number of constraints available to accomplish the desired structural elucidation.

Due to an ever expanding interest in protein structure determination, the choice of bifunctional probes for these types of substrates has dramatically increased in recent years [55–57], whereas the selection of analogous reagents for nucleic acid targets has lagged behind [58]. The majority of commercial probes are active toward sulfhydryl groups of reduced cysteines, or primary amines of lysine side-chains and N-termini. These classes of compounds exhibit little or no reactivity toward nucleobase functional groups, which precludes their application to the structural investigation of nucleic acids and protein-nucleic acid assemblies. In contrast, reagents that are active against guanido groups, such as 1,2-dicarbonyl compounds, offer the opportunity to alkylate arginine residues in proteins [59], as well as guanine nucleotides in nucleic acids [60]. In this direction, we have recently shown that β -ethoxy- α -ketobutyraldehyde (kethoxal, KT), a monofunctional 1,2-dicarbonyl employed for nucleic acid footprinting, can be successfully applied in combination with mass spectrometry to investigate the solvent accessibility of arginine side-chains in protein structures [61]. Monitoring the outcome of probing reactions by ESI on a Fourier transform ion cyclotron resonance (FTICR)[62,63] mass spectrometer, we optimized the conditions for probe application and investigated the role of borate buffer in stabilizing probe adducts by esterification of initial 1,2-diol intermediates [61]. The results clearly showed that kethoxal constitutes an excellent example of a broad-range footprinting agent for probing substrates of both proteic and nucleic acid nature.

In this report, we describe the application of homo-bifunctional crosslinkers obtained by inserting equivalent 1,2-dicarbonyl functionalities into selected aromatic scaffolds to produce 1,4-phenyl-diglyoxal (PDG, $n=1$) and 4,4'-biphenyl-diglyoxal (BDG, $n=2$) (Scheme 1). As compared to typical aliphatic scaffolds present in the vast majority of available bifunctional reagents, aromatic spacers can afford reduced flexibility and fewer rotational degrees. These

features are expected to translate into more narrowly defined crosslinking ranges that should increase the precision of the corresponding distance constraints. Additionally, the selection of phenyl and biphenyl scaffolds allowed us to test whether modular spacers of increasing size may enable closer correlations between observed crosslinks and actual distances between bridged structures. The outcome of probing reactions and the positions of crosslinked residues were interpreted in the context of the high-resolution structures of selected substrates to investigate the factors determining the formation of bridged conjugates. Matching the experimental results with the actual spatial situation of susceptible sites allowed for a fair assessment of these bifunctional probes in the 3D-structure analysis of proteins and their assemblies.

Experimental

All reagents employed in the study were obtained from commercial sources, with the exception of the bifunctional probes that were instead synthesized in house. 1,4-Diacetylbenzene, 4,4'-diacetylbiphenyl, dimethyl sulfoxide, dioxane, 48% (w/w) hydrobromic acid (HBr), and selenium dioxide (SeO₂) were purchased from Sigma-Aldrich Chemical Co. (St Louis, MO). Bovine ubiquitin (UBQ), carbonic anhydrase II (CA), trypsin, and chymotrypsin were also obtained from Sigma-Aldrich Chemical Co. and were used without further purification. Human placenta pi class glutathione S-transferase (GST) was purchased from the same supplier, but the sample was extensively desalted by ultrafiltration before use. This task was completed by using YM-3 Microcons from Millipore (Billerica, MA) to replace the original TRIS buffer with a 50 mM solution of ammonium borate adjusted to pH 8.0 [58,61]. Stock solutions in the same buffer were prepared for the model proteins with concentrations in the 100–200 μM range. The stocks were subdivided in 50 μL aliquots that were stored at 4°C until immediately before use.

The bifunctional reagents 1,4-phenyl-diglyoxal (PDG, n=1) and 4,4'-biphenyl-diglyoxal (BDG, n=2) were synthesized separately by oxidizing the respective diacetyl precursors (Scheme 1). Two alternative procedures were initially tested for PDG, which employed either SeO₂ in dioxane/water [64], or HBr in dimethyl sulfoxide [65]. Although these alternative pathways provided comparable yields in our hands, the product obtained from the former required repeated filtrations to completely remove metallic selenium produced by the redox reaction. In contrast, the latter provided a high-quality reagent that was immediately ready to use. For this reason, HBr oxidation in dimethyl sulfoxide was preferred for the synthesis of both crosslinkers. Briefly, 17 mL of 48% (w/w) HBr was added in dropwise fashion to 80 mL of dimethyl sulfoxide solutions containing 25 mmol of either 1,4-diacetylbenzene (PDB), or 4,4'-diacetyldiphenyl (BDG). The resulting solution was stirred over a 20 min period and then heated to 60°C for 18 h. After the reaction mixture had been allowed to cool, it was poured into 600 mL of ice-water. The precipitate was filtered, washed with cold water, and then dried under suction. The correct identity of the desired products was confirmed by mass spectrometry and NMR spectroscopy [66]. Due to the absence of readily ionizable groups in the crosslinkers structures, their MS analysis was completed only after forming adducts with N-acetylarginine, as shown earlier for kethoxal [61]. Stock solutions containing 5–10 mM of PDG and BDG were prepared in dimethyl-formamide (DMF) and stored at –20°C until immediately before use.

Typical 25 μL reactions were setup by mixing appropriate volumes of substrate and crosslinker stocks in 50 mM ammonium borate buffer (pH adjusted to 8.0). For each substrate, the reactivity of the different crosslinkers was initially tested using a 1:1 probe-to-substrate ratio (~50 μM each), but ratios ranging up to 20:1 were subsequently explored. Initial experiments were completed by performing reactions at 4, 25, 37, and 45°C, and by monitoring the results as a function of time for up to 12 h. After considering the preliminary results, reactions were typically carried out at 37°C for 1.5 h, using a 4:1 probe-to-substrate ratio for UBQ, 9:1 for

CA, and 16:1 for GST. At the end of the reaction interval, unreacted probe was removed by dilution with buffer and ultrafiltration on YM-3 Microcons. If necessary, sample solutions were stored at -20°C and thawed immediately prior to MS analysis or protease digestion. A $1\ \mu\text{L}$ aliquot containing $1\ \text{mg mL}^{-1}$ of either trypsin or chymotrypsin in $10\ \text{mM}$ ammonium borate ($\text{pH } 8.0$) was added to a $25\ \mu\text{L}$ crosslinking sample to generate the desired peptide maps. Unless otherwise indicated, digestion was performed at 37°C overnight.

All analyses were performed on a Bruker Daltonics (Billerica, MA) Apex III FTICR mass spectrometer equipped with a 7T actively-shielded superconducting magnet and a nano-ESI source built in house. Each sample was diluted to a final concentration of approximately $5\ \mu\text{M}$ using either water/methanol/acetic acid (49:49:2 in volume), or $10\ \text{mM}$ ammonium borate ($\text{pH } 8.0$). Typically, $5\ \mu\text{L}$ samples were loaded into the nano-ESI needle and a spray voltage of less than $1\ \text{kV}$ was applied to the solution through a stainless steel wire inserted in the needle's back end. No solvent pumps were necessary, as the solution flow-rate was dictated by the applied voltage and the size of the nano-ESI needle tip (typically $\sim 1\text{--}2\ \mu\text{m}$). Spectra were acquired in positive ionization mode and processed using XMASS 7.0.2 (Bruker Daltonics, Billerica, MA). Scans were completed in broadband mode that afforded a typical $150,000$ resolving power at m/z 2000. The spectra were externally calibrated using a $1\ \text{mg mL}^{-1}$ solution of CsI, which produced a series of peaks throughout the mass range of $1000\text{--}6000\ m/z$ and enabled to achieve a typical mass accuracy of $20\ \text{ppm}$ or better across the range. Each analysis was performed a minimum of three times. Theoretical mass lists of digestion products were generated using the PeptideMass software available in the ExpASY suite [67], or the LINKS algorithm that was specifically developed to aid the characterization of crosslinked biomolecules ([68] and refs. therein).

Results and Discussion

The presence of four arginine residues in its sequence makes ubiquitin (UBQ) an ideal substrate for testing Arg-specific probes, as demonstrated by our earlier study of the alkylating properties of the monofunctional 1,2-dicarbonyl kethoxal [61]. A rather compact fold comprising a short α -helix and a five-stranded β -sheet with a highly hydrophobic interface [69] is expected to confer different levels of accessibility to the potential modification sites R42, R54, R72, and R74 (Table 1). Additionally, its relatively small size (76 amino acids, 8559.62 Da monoisotopic mass) enables the application of ESI-FTICR to analyze intact substrates and to assess the extent of modification without proteolytic digestion. Taking advantage of this favorable feature, a series of alkylation products were readily detected in reaction mixtures obtained by treating UBQ with either PDG or BDG in $50\ \text{mM}$ ammonium borate buffer ($\text{pH } 8.0$) for 1.5 h at 37°C (Figure 1, see *Material and Methods*). For both reagents, up to four adducts were observed with characteristic series of products consistent with the known reaction mechanism and with the formation of borate diesters involving any of the initial 1,2-diol species (*e.g.*, **B1**, **B2**, and **B3** in Scheme 2) [58,61]. The maximum number of adducts did not change by increasing probe-to-substrate ratio or reaction time, which confirmed the specificity of 1,2-dicarbonyl activity toward arginine side-chains [61]. As shown in Figure 1, no unmodified substrate could be detected after 1.5 h incubation at 37°C with a 4:1 probe-to-substrate ratio of either PDG or BDG. Similar outcome with no unmodified UBQ left in solution was achieved by using a 12:1 ratio of the kethoxal analogue under otherwise identical conditions [61]. This observation indicates that the electron-withdrawing properties of aromatic substituents increase significantly the reactivity of 1,2-dicarbonyls under the selected experimental conditions.

The nature of the reaction between 1,2-dicarbonyl and guanido groups is such that mono- and bifunctional adducts share identical incremental mass (see for example **A1** and **A2** in Scheme 2), which may complicate data interpretation. As exemplified by PDG mono-adducts displayed in Figure 2b, the species observed with a monoisotopic mass of $8731.57\ \text{Da}$ could correspond

to either of the isobaric products **A1-H₂O** and **A2-H₂O**, which are expected to have a mass of 8731.64 Da calculated from sequence using $\Delta m = 172.02$ Da (Scheme 2). For the same reason, the species with experimental mass 8749.58 and 8775.57 Da could match respectively **A1/A2** and **B1/B2** (9749.65 and 8775.65 Da calc., $\Delta m = 190.03$ and 216.03 Da). In contrast, the species with experimental mass 8757.57 Da corresponds exclusively to **B2-H₂O** (8757.64 Da calc., $\Delta m = 198.02$ Da). According to the putative structures in Scheme 2, only bifunctional products afford two distinct 1,2-diol functions that are capable of producing a borate diester and formally losing water at the same time. Therefore, **B2-H₂O** offers unambiguous evidence that the mono-adduct population in Figure 2b included bifunctional crosslinks. The detection of unique diagnostic products becomes critical for the interpretation of data obtained from substrates capable of supporting multiple modifications, as exemplified by the PDG bis-adducts in Figure 2c, which could be assigned to different combinations of isobaric products. In this region, the most intense species *e* matches the combination of one hydrated monofunctional product **A3** with the borate diester of either a mono- or a bifunctional adduct (**B1/B2**). Species *b*, instead, can only correspond to the combination of **B1/B2** with the unique **A2-2H₂O** product. According to Scheme 2, both mono- and bifunctional adducts can formally lose *one* water molecule, but only the latter can eliminate *two* to produce **A2-2H₂O**. Therefore, characteristic incremental masses of 154.01 and 230.04 Da constitute unambiguous diagnostics for the formation of desirable crosslinked conjugates with PDG and BDG, respectively.

The position of modified residues was characterized using a classic bottom-up strategy involving the cleavage of probed substrates by different proteases to yield overlapping maps. Chymotrypsin was selected as a complement to trypsin for overcoming a possible decrease of tryptic activity caused by arginine modifications [61]. Applied to PDG and BDG reaction mixtures, each protease provided series of hydrolytic peptides, which afforded excellent sequence coverage and allowed for the characterization of probe adducts. The identification of conjugated products was completed by using LINKS [68], a software package designed to support the analysis of crosslinked biomolecules. As summarized in Table 2, the results indicated that the crosslinkers were capable of modifying all available sites, although with different outcomes. The observed alkylation patterns were consistent with the estimate of solvent accessibility indexes obtained by applying the GETAREA [70] algorithm to the UBQ crystal structure (PDB ID: 1ARR) [69], which provided a 40% index for R42, 55% for R54, 54% for R72, and 85% for R74. According to this algorithm, any residue with an index greater than 50% is considered exposed, whereas those below 20% are described as buried. The fact that R42 exhibited the lowest index in the series may account for the low abundance recorded for the corresponding modifications.

The modification patterns provided by PDG and BDG were examined in the context of the structural situation of each modified residue to investigate the factors that influence the formation of sought-after bridging products. A close look at the crystal structure reveals that R42, R72, and R74 are placed within close proximity by the protein fold, whereas R54 is isolated in a distal region (Figure 3). Considering that PDG and BDG possess crosslinking ranges of 6.14 ± 0.64 Å and 10.44 ± 0.80 Å, respectively, the absence of bifunctional products involving R54 can be clearly explained by its placement ~ 17.9 Å away from the closest susceptible residue, R42. Although the high-resolution structure displays a distance of only ~ 4.8 Å between the guanido groups of contiguous R72 and R74, these sites are capable of stretching farther apart due to inherent flexibility and rotational freedom typical of arginine side-chains. The fact that only the shorter reagent PDG produced a detectable bifunctional conjugate (Table 2) indicates that these side-chains cannot extend far enough to match the effective range of BDG. This notable example highlights a glaring deficiency in how residues spacing is generally determined using graphics software that treats high-resolution structures as rather rigid and static entities. Each spacing distance should include a range that accounts for intrinsic flexibility and rotational effects typical of side-chains that are not locked in specific

conformations [71]. Only this type of treatment would be expected to enable meaningful matching between crosslinking ranges and residues spacing.

Based on traditional distance measurements alone, it was possible to predict the formation of bifunctional crosslinks between R42 and either R72 or R74, which are placed on average ~ 8.0 Å from each other (Figure 3). The absence of conjugated products from either reagent could stem not only from the reduced level of accessibility manifested by R42, but also from the relatively unconstrained nature of the C-terminal tail containing R72 and R74. In fact, while the reference crystal structure places these residues within crosslinking range, numerous NMR structures in the Protein Data Bank show that the C-terminal region can assume widely different conformations. It is possible that the crosslinkable form may represent a rare conformation in solution, which is instead favored by crystal packing effects. If this is the case, the fact that an extended time-course analysis failed to detect R42-R72/R74 conjugates should be interpreted as a sign that such residues do not spend enough time within crosslinking range to allow for bridging to take place in solution.

The results provided by UBQ suggest that solvent accessibility and proper spacing are critical, but not sufficient to explain the formation of bridged products. The significance of putative dynamic effects in determining the outcome of crosslinking reactions can be better evaluated in a more complex substrate, such as bovine carbonic anhydrase II (CA). The sequence of this protein consists of 259 amino acids (28964.67 Da monoisotopic mass) and includes 9 arginine residues (Table 1). Upon structural probing and protease digestion, ESI-FTICR analysis afforded peptide maps that unambiguously identified the position of modified amino acids (Table 3). The presence of favorable proteolytic sites distributed between alkylated arginines resulted in the formation of conjugated products consisting of distal peptides bridged by the crosslinker structures, which were immediately recognizable from their unique molecular masses. When no cleavage sites are available between two modified residues, as exemplified by the R72-R74 conjugate observed in UBQ (Table 2), any intra-peptide crosslink can be discriminated from the corresponding monofunctional adduct only by the characteristic incremental masses of 154.01 and 230.04 Da discussed above.

The majority of the probed products obtained from CA were consistent with the accessibility and structural context of the amino acids involved, including PDG crosslinking of R56-R57 and R26-R255. According to the high-resolution structure (PDB ID: 1V9E)[72], these residues possess the following solvent accessibility indexes: 58% for R56; 49% for R57; 38% for R26; 29% for R255. In analogy with the R72-R74 conjugate observed in UBQ, the R56-R57 involves contiguous residues that are instead extended ~ 11.5 Å apart in the crystal structure, but are capable of rotating closer to one another due to their unconstrained location on the protein surface. In the case of R26-R255, the susceptible functional groups are spaced by ~ 13.94 Å on the crystal structure, but R255 is placed near the C-terminal end, where it is likely subjected to significant dynamics in solution.

A spacing of ~ 13.98 Å and accessibility indexes of 58% and 16% cannot immediately justify the formation of a conjugate between R56 and R88 because of the presence of an interposed loop that appears to block the line of sight between their respective guanido groups (Figure 4). In this case, the formation of R56-R88 is possible only if the arginine side-chains assume alternative conformations and the intervening loop enjoys a relatively high level of flexibility. This observation highlights an important issue faced when interpreting the results of chemical probing, which can be affected by local dynamics and may reflect the kinetic trapping of sparsely populated conformational states. Although a detailed discussion is beyond the scope of this report, these types of situations can be readily recognized by monitoring the outcome of probing reactions performed after varying probe-to-substrate ratio, temperature, time, pH, salt content, and any environmental condition that may influence substrate conformation.

Parallel probing with complementary reagents is expected to afford corroborating information necessary to accurately identify regions involved in significant dynamics in solution. In this direction, the fact that both PDG and BDG were capable of bridging R56 and R88 despite their different spacers (Table 3) indicates that these residues can assume widely different conformations in 3D space. The concerted application of nested reagents could provide a very effective tool for identifying flexible domains and investigating structure dynamics.

Bifunctional reagents can afford valuable information about the organization of multi-subunit assemblies by crosslinking residues at the interface between cognate components. The ability of 1,2-dicarbonyl reagents to form inter-subunit conjugates was tested on human placenta pi class glutathione S-transferase (GST). This dimeric substrate includes two identical subunits, each consisting of 209 amino acids (23209.98 Da monoisotopic mass, Table 1). Solvent accessibility estimated from the high-resolution structure (PDB ID: 14GS)[73] suggested that only half of the eight arginines on each subunit are sufficiently exposed to serve as potential targets. Indeed, GETAREA provided the following accessibility indexes: 32% for R182 and R186; 27% for R13; 17% for R74; 13% for R11; 5% for R18 and R100; 2% for R70. In each monomeric structure, the pairs R13-R74 and R182-R186 are located within crosslinkable distance, but out of range from each other. Considering a ~ 16.77 and ~ 11.02 Å spacing for the R13-R74 and R182-R186 pairs, respectively, and allowing for side-chains rotational freedom, it was not surprising that crosslinks were readily observed with the longer probe BDG but not with the shorter reagent PDG that produced only mono-functional adducts (Table 4). In addition, a series of bridged peptides were detected, revealing the crosslinking of R13 and R74 with their equivalent counterpart on the other subunit of the dimeric complex (*i.e.*, R13(a)-R13(b) and R74(a)-R74(b)). A close examination of the high-resolution structure shows that corresponding R13 residues point toward each other across the dimer interface, assuming ideal positions for inter-molecular crosslinking (Figure 5). Favorable positioning is observed also for cognate R74 residues of bound subunits, which supports the formation of their conjugated product. The different spacing observed for the two pairs explains why R74(a)-R74(b) with susceptible functional groups placed 7.36 Å from each other was produced only by the shorter PDG, whereas R13(a)-R13(b) spaced by ~ 13.54 Å required the longer crosslinker BDG. Had the dimeric structure of GST been unknown, these conjugates would have revealed unambiguously the proper placement and reciprocal orientation of bound monomers, with each R13 facing the equivalent residue across the dimer interface, rather than the R74 of the cognate subunit. This type of information cannot be readily obtained from monofunctional approaches (*e.g.*, footprinting and HD exchange), which can only reveal whether a certain residue is protected, but cannot immediately identify the source of the observed protection effects.

Conclusions

The results provided by PDG and BDG demonstrate the ability of *bis*-(1,2-dicarbonyls) to induce bifunctional crosslinking of Arg-containing proteins. The introduction of aromatic spacers in place of typical aliphatic structures had significant effects on the overall reactivity towards guanido groups, which was greater than that of the aliphatic monofunctional reagent kethoxal [61]. Efficient probing was achieved under relatively mild conditions of pH and ionic strength, which are expected to preserve the normal folding of target biomolecules. As observed for other 1,2-dicarbonyl adducts with protein and nucleic acid substrates [58,61], the presence of borate in the solution can stabilize initial 1,2-diol products by producing the respective diesters. The formation of multiple related products and the fact that mono- and bifunctional adducts share the same incremental mass tend to complicate the immediate characterization of bridged products from direct molecular mass determinations. However, the application of appropriate proteolytic procedures to induce cleavage between crosslinked sites can yield conjugated digestion peptides with unique molecular masses that allow for unambiguous identification. The application of top-down techniques in which sequencing by tandem mass

spectrometry is performed directly on the crosslinked substrates could enable the characterization of conjugated residues with no need for a proteolytic step.

In planning the application of these bifunctional reagents, the actual presence of Arg residues in the target sequence, their accessibility on the substrate surface, a proper match between inter-residue distances and crosslinker span, and the presence of a clear path between susceptible sites are important factors that may determine the outcome of probing reactions. Some of these factors can be immediately evaluated when either a high-resolution structure or a working model is available, with the caveat that any prediction of inter-residue distances should always account for the flexibility and rotational freedom of targeted side-chains in their actual structural context [71]. Conversely, each distance constraint employed for modeling unknown substrates should include reasonable ranges that factor in appropriate statistical information on both probe and substrate flexibility. In this direction, the probing community is striving to refine the way these types of spatial constraints should be evaluated and used in modeling operations.

In tackling unknown substrates, proper experimental design is paramount for recognizing and evaluating the incidence of possible structure dynamics. Probe-to-substrate ratios should always be kept as low as possible to reduce crosslinking yields and to minimize the risk of detecting sparsely populated conformational states [25,74]. Titration schemes in which the amount of probe is increased in stepwise fashion can effectively differentiate primary modification sites from secondary sites that are exposed only as a result of conformational changes induced by initial alkylation [68]. Yield and position of crosslinked products should be also evaluated by repeating probing reactions under a wide variety of environmental conditions that influence substrate conformation [25]. Finally, complementary probes with different target specificity and crosslinking span should be employed in concert to check for diagnostic inconsistencies. In this work, we have clearly shown that the direct correlation between elevated structural flexibility and comparable results by PDG and BDG can provide a powerful tool for identifying highly dynamic regions.

Thanks to their specificity for Arg residues, PDG and BDG constitute excellent complements to typical Cys- and Lys-specific crosslinkers available on the market. The observation that the majority of known RNA-binding proteins contain characteristic arginine-rich motifs [75] makes these reagents particularly appealing for the investigation of ribonucleoprotein assemblies. In light of the ability of 1,2-dicarbonyls to form stable adducts with the base-pairing face of guanine [58], these reagents afford the possibility of producing protein-protein, protein-nucleic acids, and nucleic acids-nucleic acids conjugates in the same experiment, thus maximizing the information attainable from individual probing reactions. The possibility of employing series of nested reagents in multiplexed schemes should also help increasing the number of experimental constraints available for modeling [76]. The fact that these types of conjugates can be readily characterized by ESI-FTICR constitutes a definite strength of this concerted approach for structural elucidation. The broad applicability of this analytical platform and the ongoing development of top-down strategies for the characterization of crosslinked species are expected to support the structural analysis of biomolecular assemblies of increasing size and complexity. This type of endeavor, however, will greatly benefit from expanding the selection of bifunctional reagents for targeting a wider variety of functional groups and for obtaining a larger number of valid spatial constraints to complete the actual modeling of target substrates.

Acknowledgements

This research was funded by the National Institutes of Health (GM643208) and by the National Science Foundation (CHE-0439067).

References

1. Chowdhury SK, Katta V, Chait BT. Probing conformational changes in proteins by mass spectrometry. *J Am Chem Soc* 1990;112:9012–9012.
2. Wales TE, Engen JR. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrom Rev* 2006;25:158–170. [PubMed: 16208684]
3. Bennett KL, Matthiesen T, Roepstorff P. Probing protein surface topology by chemical surface labeling, crosslinking, and mass spectrometry. *Methods Mol Biol* 2000;146:113–31. [PubMed: 10948499]
4. Guan JQ, Chance MR. Structural proteomics of macromolecular assemblies using oxidative footprinting and mass spectrometry. *Trends Biochem Sci* 2005;30:583–92. [PubMed: 16126388]
5. Peattie DA, Gilbert W. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci USA* 1980;77:4679–82. [PubMed: 6159633]
6. Matthews KS, Chakerian AE, Gardner JA. Protein chemical modification as probe of structure-function relationships. *Methods Enzymol* 1991;208:468–96. [PubMed: 1779844]
7. Brimacombe R, Stiege W, Kyriatsoulis A, Maly P. Intra-RNA and RNA-protein crosslinking techniques in *Escherichia coli* ribosomes. *Methods Enzymol* 1988;164:287–309. [PubMed: 3071669]
8. Expert-Bezançon A, Chiaruttini C. RNA-protein crosslinking. *Methods Enzymol* 1988;164:310–318. [PubMed: 3071671]
9. Gaffney BJ. Chemical and biochemical crosslinking of membrane components. *Biochim Biophys Acta* 1985;822:289–317. [PubMed: 2865974]
10. Tanaka K, Waki H, Ido H, Akita S, Yoshida T. Protein and polymer analysis up to 100,000 by laser ionization time-of-flight mass spectrometry. *Rapid Commun Mass Spectrom* 1988;2:151–153.
11. Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 Daltons. *Anal Chem* 1988;60:2299–2301. [PubMed: 3239801]
12. Yamashita M, Fenn JB. Electrospray ion source. Another variation on the free-jet theme. *J Phys Chem* 1984;88:4671–4675.
13. Aleksandrov ML, Gall LN, Krasnov VN, Nikolaev VI, Pavlenko VA, Shkurov VA. Extraction of ions from solutions under atmospheric pressure: a method of mass spectrometric analysis of bioorganic compounds. *Doklady Akademii Nauk* 1984;277:379–383.
14. Hunt DF, Shabanowitz J, Yates JRr, Zhu NZ, Russell DH, Castro ME. Tandem quadrupole Fourier transform mass spectrometry of oligopeptides and small proteins. *Proc Natl Acad Sci USA* 1987;84:60–63.
15. Biemann K, Scoble H. Characterization by tandem mass spectrometry of structural modifications in proteins. *Science* 1987;237:992–998. [PubMed: 3303336]
16. Haniu M, Narhi LO, Arakawa T, Elliott S, Rohde MF. Recombinant human erythropoietin (rHuEPO): cross-linking with disuccinimidyl esters and identification of the interfacing domains in EPO. *Protein Sci* 1993;2:1441–51. [PubMed: 8401229]
17. Lacroix M, Rossi V, Gaboriaud C, Chevallier S, Jaquinod M, Thielens NM, Gagnon J, Arlaud GJ. Structure and assembly of the catalytic region of human complement protease C1r: a three-dimensional model based on chemical cross-linking and homology modeling. *Biochemistry* 1997;36:6270–82. [PubMed: 9174342]
18. Sharma KK, Kaur H, Kester K. Functional elements in molecular chaperone alpha-crystallin: identification of binding sites in alpha B-crystallin. *Biochem Biophys Res Commun* 1997;239:217–22. [PubMed: 9345298]
19. Young MM, Tang N, Hempel JC, Oshiro CM, Taylor EW, Kuntz ID, Gibson BW, Dollinger G. High throughput protein fold identification by using experimental constraints derived from intramolecular cross-links and mass spectrometry. *Proc Natl Acad Sci U S A* 2000;97:5802–6. [PubMed: 10811876]
20. Pearson KM, Pannell LK, Fales HM. Intramolecular Cross-Linking Experiments on Cytochrome C and Ribonuclease a Using an Isotope Multiplet Method. *Rapid Commun Mass Spectrom* 2002;16:149–159. [PubMed: 11803535]
21. Kruppa G, Schoeniger J, Young MC. A top down approach to protein structural studies using chemical cross-linking and Fourier transform mass spectrometry. *Rapid Commun Mass Spectrom* 2003;17:155–162. [PubMed: 12512095]

22. Novak P, Young M, Schoeniger J, Kruppa GH. A top down approach to protein structure studies using chemical cross-linking and Fourier transform mass spectrometry. *Eur J Mass Spectrom* 2003;9:623–31.
23. Peterson JJ, Young MM, Takemoto LJ. Probing alpha-crystallin structure using chemical cross-linkers and mass spectrometry. *Mol Vis* 2004;10:857–66. [PubMed: 15570221]
24. Silva RA, Hilliard GM, Fang J, Macha S, Davidson WS. A three-dimensional molecular model of lipid-free apolipoprotein A-I determined by cross-linking/mass spectrometry and sequence threading. *Biochemistry* 2005;44:2759–69. [PubMed: 15723520]
25. Yu ET, Zhang Q, Fabris D. Untying the FIV frameshifting pseudoknot structure by MS3D. *J Mol Biol* 2005;345:69–80. [PubMed: 15567411]
26. Yu ET, Hawkins AE, Eaton J, Fabris D. MS3D structural elucidation of the HIV-1 packaging signal. 2008Submitted
27. Schulz DM, Ihling C, Clore GM, Sinz A. Mapping the topology and determination of a low-resolution three-dimensional structure of the calmodulin-melittin complex by chemical cross-linking and high-resolution FTICRMS: direct demonstration of multiple binding modes. *Biochemistry* 2004;43:4703–15. [PubMed: 15096039]
28. Kalkhof S, Ihling C, Mechtler K, Sinz A. Chemical cross-linking and high-performance Fourier transform ion cyclotron resonance mass spectrometry for protein interaction analysis: application to a calmodulin/target peptide complex. *Anal Chem* 2005;77:495–503. [PubMed: 15649045]
29. Pohl T, Wittmann-Liebold B. Identification of a cross-link in the Escherichia coli ribosomal protein pair S13–S19 at the amino acid level. *J Biol Chem* 1988;263:4293–301. [PubMed: 3279034]
30. Farmer TB, Caprioli RM. Assessing the multimeric states of proteins: studies using laser desorption mass spectrometry. *Biol Mass Spectrom* 1991;20:796–800. [PubMed: 1812990]
31. Yang T, Horejsh DR, Mahan KJ, Zaluzec EJ, Watson TJ, Gage DA. Mapping cross-linking sites in modified proteins with mass spectrometry: an application to cross-linked hemoglobins. *Anal Biochem* 1996;242:55–63. [PubMed: 8923964]
32. Yu Z, Friso G, Miranda JJ, Patel MJ, Lo-Tseng T, Moore EG, Burlingame AL. Structural characterization of human hemoglobin crosslinked by bis(3,5-dibromosalicyl) fumarate using mass spectrometric techniques. *Protein Sci* 1997;6:2568–77. [PubMed: 9416606]
33. Borchers C, Tomer KB. Characterization of the noncovalent complex of human immunodeficiency virus glycoprotein 120 with its cellular receptor CD4 by matrix-assisted laser desorption/ionization mass spectrometry. *Biochemistry* 1999;38:11734–40. [PubMed: 10512629]
34. McLachlin DT, Coveny AM, Clark SM, Dunn SD. Site-directed cross-linking of b to the alpha, beta, and a subunits of the Escherichia coli ATP synthase. *J Biol Chem* 2000;275.
35. Rappsilber J, Siniossoglou S, Hurt EC, Mann M. A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Anal Chem* 2000;72:267–75. [PubMed: 10658319]
36. Back JW, Hartog AF, Dekker HL, Muijsers AO, de Koning LJ, de Jong L. A new crosslinker for mass spectrometric analysis of the quaternary structure of protein complexes. *J Am Soc Mass Spectrom* 2001;12:222–7. [PubMed: 11212007]
37. Sinz A, Wang K. Mapping protein interfaces with a fluorogenic cross-linker and mass spectrometry: application to nebulin-calmodulin complexes. *Biochemistry* 2001;40:7903–13. [PubMed: 11425319]
38. Wine RN, Dial JM, Tomer KB, Borchers CH. Identification of components of protein complexes using a fluorescent photo-cross-linker and mass spectrometry. *Anal Chem* 2002;74:1939–45. [PubMed: 12033289]
39. Chang Z, Kuchar J, Hausinger RP. Chemical cross-linking and mass spectrometric identification of sites of interaction for UreD, UreF, and urease. *J Biol Chem* 2004;279:15305–13. [PubMed: 14749331]
40. Carlsohn E, Ångström J, Emmet MR, Marshall AG, Nilsson CL. Chemical crosslinking of the urease complex from *Helicobacter pylori* and analysis by Fourier transform ion cyclotron resonance mass spectrometry and molecular modeling. *Int J Mass Spectrom* 2004;234:137–144.
41. Gingras AC, Gstaiger M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. *Nat Rev Mol Cell Biol* 2007;8:645–54. [PubMed: 17593931]

42. Urlaub H, Krufft V, Bischof O, Muller EC, Wittmann-Liebold B. Protein-rRNA binding features and their structural and functional implications in ribosomes as determined by cross-linking studies. *Embo J* 1995;14:4578–88. [PubMed: 7556101]
43. Sastry SS. Identification of the template-binding cleft of T7 RNA polymerase as the site for promoter binding by photochemical cross-linking with psoralen. *Biochemistry* 1996;35:13519–30. [PubMed: 8885831]
44. Jensen ON, Kulkarni S, Aldrich JV, Barofsky DF. Characterization of peptide-oligonucleotide heteroconjugates by mass spectrometry. *Nucleic Acids Res* 1996;24:3866–72. [PubMed: 8871569]
45. Urlaub H, Thiede B, Muller EC, Brimacombe R, Wittmann-Liebold B. Identification and sequence analysis of contact sites between ribosomal proteins and rRNA in *Escherichia coli* 30 S subunits by a new approach using matrix-assisted laser desorption/ionization-mass spectrometry combined with N-terminal microsequencing. *J Biol Chem* 1997;272:14547–55. [PubMed: 9169412]
46. Urlaub H, Thiede B, Muller EC, Wittmann-Liebold B. Contact sites of peptide-oligoribonucleotide cross-links identified by a combination of peptide and nucleotide sequencing with MALDI MS. *J Protein Chem* 1997;16:375–83. [PubMed: 9246617]
47. Naryshkin NA, Farrow MA, Ivanovskaya MG, Oretskaya TS, Shabarova ZA, Gait MJ. Chemical cross-linking of the human immunodeficiency virus type 1 Tat protein to synthetic models of the RNA recognition sequence TAR containing site-specific trisubstituted pyrophosphate analogues. *Biochemistry* 1997;36:3496–505. [PubMed: 9131999]
48. Wong DL, Pavlovich JG, Reich NO. Electrospray ionization mass spectrometric characterization of photocrosslinked DNA-EcoRI DNA methyltransferase complexes. *Nucleic Acids Res* 1998;26:645–9. [PubMed: 9421528]
49. Golden MC, Resing KA, Collins BD, Willis MC, Koch TH. Mass spectral characterization of a protein-nucleic acid photocrosslink. *Protein Sci* 1999;8:2806–12. [PubMed: 10631998]
50. Urlaub H, Hartmuth K, Kostka S, Grelle G, Luhrmann R. A general approach for identification of RNA-protein cross-linking sites within native human spliceosomal small nuclear ribonucleoproteins (snRNPs). Analysis of RNA-protein contacts in native U1 and U4/U6.U5 snRNPs. *J Biol Chem* 2000;275:41458–68. [PubMed: 11006293]
51. Steen H, Jensen ON. Analysis of protein-nucleic acid interactions by photochemical cross-linking and mass spectrometry. *Mass Spectrom Rev* 2002;21:163–82. [PubMed: 12476441]
52. Steen H, Petersen J, Mann M, Jensen ON. Mass spectrometric analysis of a UV-cross-linked protein-DNA complex: tryptophans 54 and 88 of *E. coli* SSB cross-link to DNA. *Protein Sci* 2001;10:1989–2001. [PubMed: 11567090]
53. Urlaub H, Hartmuth K, Luhrmann R. A two-tracked approach to analyze RNA-protein crosslinking sites in native, nonlabeled small nuclear ribonucleoprotein particles. *Methods* 2002;26:170–81. [PubMed: 12054894]
54. Geyer H, Geyer R, Pingoud V. A novel strategy for the identification of protein-DNA contacts by photocrosslinking and mass spectrometry. *Nucl Acids Res* 2004;32:e132. [PubMed: 15383647]
55. Pierce applications handbook and catalog. Pierce Biotechnology, Inc.; Rockford IL: 2005.
56. Back JW, de Jong L, Muijsers AO, de Koster CG. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* 2003;331:303–313. [PubMed: 12888339]
57. Sinz A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J Mass Spectrom* 2003;38:1225–1237. [PubMed: 14696200]
58. Zhang Q, Yu ET, Kellersberger KA, Crosland E, Fabris D. Toward building a database of bifunctional probes for the MS3D investigation of nucleic acids structures. *J Am Soc Mass Spectrom* 2006;17:1570–1581. [PubMed: 16875836]
59. Iijima H, Patrzyk H, Bello J. Modification of amino acids and bovine pancreatic ribonuclease A by kethoxal. *Biochim Biophys Acta* 1977;491:305–316. [PubMed: 14699]
60. Shapiro R, Cohen BI, Shiuuey SJ, Maurer H. On the reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. A new synthesis of N2-alkylguanines. *Biochemistry* 1969;8:238–45. [PubMed: 5777326]
61. Akinsiku OT, Yu ET, Fabris D. Mass spectrometric investigation of protein alkylation by the RNA footprinting probe kethoxal. *J Mass Spectrom* 2005;40:1372–1381. [PubMed: 16237662]

62. Comisarow MB, Marshall AG. Fourier transform ion cyclotron resonance. *Chem Phys Lett* 1974;25:282–283.
63. Hendrickson CL, Emmett MR, Marshall AG. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Annu Rev Phys Chem* 1999;50:517–536. [PubMed: 10575730]
64. Wagner R, Garrett RA. A new RNA-RNA crosslinking reagent and its application to ribosomal 5S RNA. *Nucleic Acids Res* 1978;5:4065–75. [PubMed: 724507]
65. Bruce MJ, McLean GA, Royles BJL, Smith DM, Stranding PN. The ‘inverse electron-demand’ Diels–Alder reaction in polymer synthesis. Part 2 Some bis (1, 2, 4-triazines) as potential bis-diene monomers. *J Chem Soc Perkin Trans* 1995;1:1789–95.
66. NMR analysis was completed on a 400 MHz JEOL (Tokyo, Japan) ECX400 instrument. PDG: ^1H NMR (400 MHz, DMSO-D_6) $\delta=9.47$ (d, $^5\text{J}=2.3$ Hz, 1H, CHO), 8.13 (dd, $^3\text{J}=16$ Hz, $^5\text{J}=4.6$ Hz, 4H, PhCH₄) ppm. BDG: ^1H NMR (400 MHz, DMSO-D_6) $\delta=9.56$ (d, $^5\text{J}=11.0$ Hz, 1H, CHO), 8.20 (s, 4H, H-3, H-5, H-3', H-5'), 7.94 (dd, $^3\text{J}=9.6$ Hz, $^5\text{J}=8.2$, 4H, H-2, H-6, H-2', H-6') ppm.
67. Wilkins MR, Lindskog I, Gasteiger E, Bairoch A, Sanchez JC, Hochstrasser DF, Appel RD. Detailed peptide characterisation using PEPTIDEMASS - a World-Wide Web accessible tool. *Electrophoresis* 1997;18:403–408. [PubMed: 9150918]
68. Kellersberger KA, Yu E, Kruppa GH, Young MM, Fabris D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal Chem* 2004;76:2438–2445. [PubMed: 15117181]
69. Cook WJ, Jeffrey LC, Carson M, Chen Z, Pickart CM. Structure of a diubiquitin conjugate and a model for interaction with ubiquitin conjugating enzyme (E2). *J Biol Chem* 1992;267:16467–16471. [PubMed: 1322903]
70. Franczkiewicz R, Braun W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J Comp Chem* 1998;19:319–333.
71. Schoeniger, J. Protein structure and conformation revealed using sequential cross-linker rate measurements (SCRaM): direct measurement of inter-residue distances and secondary structure. Proceedings of the 55th American Society for Mass Spectrometry Conference; June 3–7, 2007; Indianapolis, IN.
72. Saito R, Sato T, Ikai A, Tanaka N. Structure of bovine carbonic anhydrase II at 1.95 Å resolution. *Acta Crystallogr, Sect D* 2004;60:792–795. [PubMed: 15039588]
73. Oakley AJ, Lo Bello M, Ricci G, Federici G, Parker MW. Evidence for an induced-fit mechanism operating in pi class glutathione transferases. *Biochemistry* 1998;37:9912–9917. [PubMed: 9665696]
74. Yu E, Fabris D. Direct probing of RNA structures and RNA-protein interactions in the HIV-1 packaging signal by chemical modification and electrospray ionization Fourier transform mass spectrometry. *J Mol Biol* 2003;330:211–223. [PubMed: 12823962]
75. Weiss MA, Narayana N. RNA recognition by arginine-rich peptide motifs. *Biopolymers* 1998;48:167–80. [PubMed: 10333744]
76. Yu ET, Fabris D. Toward multiplexing the application of solvent accessibility probes for the investigation of RNA three-dimensional structures by electrospray ionization - Fourier transform mass spectrometry. *Anal Biochem* 2004;344:356–366. [PubMed: 15494143]

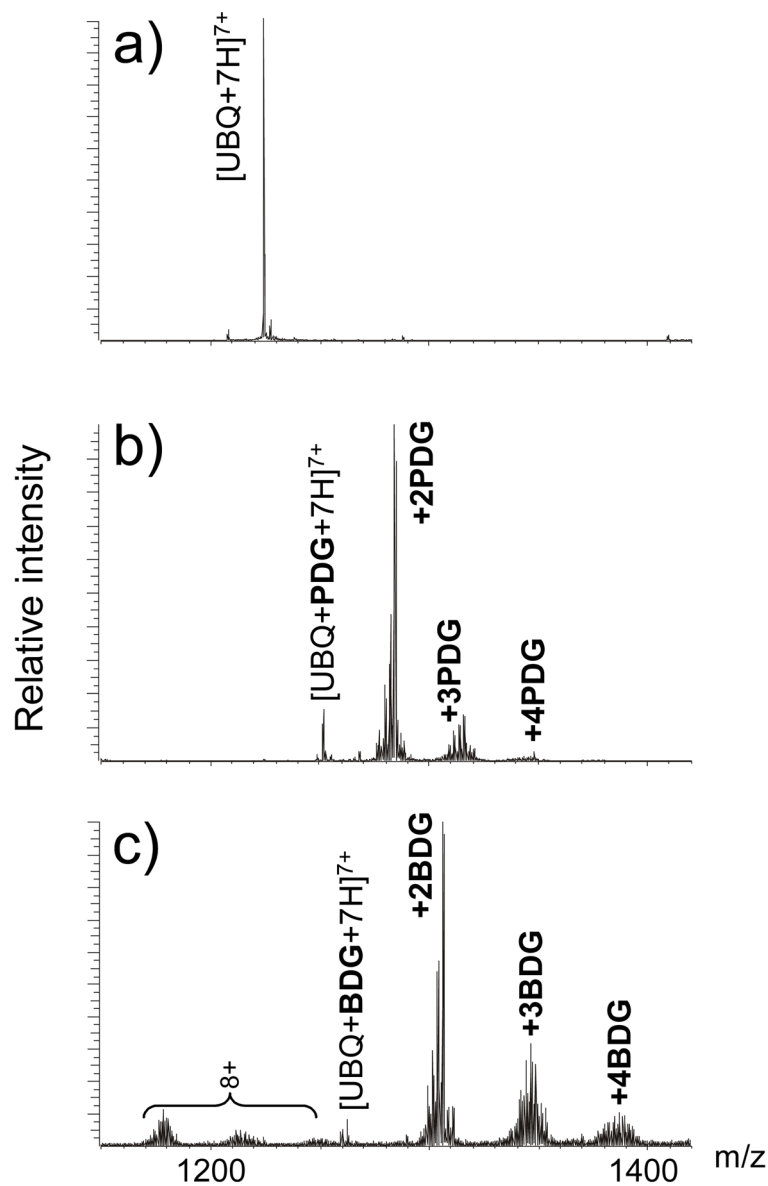


Figure 1. Representative ESI-FTICR spectra of ubiquitin (UBQ) before **a)** and after reaction with **b)** 1,4-phenyl-diglyoxal (PDG, n=1) and **c)** 4,4'-biphenyl-diglyoxal (BDG, n=2). The regions containing the 7+ charge state are shown. These results were obtained by incubating the substrate for 1.5 h at 37°C in the presence of a 4:1 probe-to-substrate ratio.

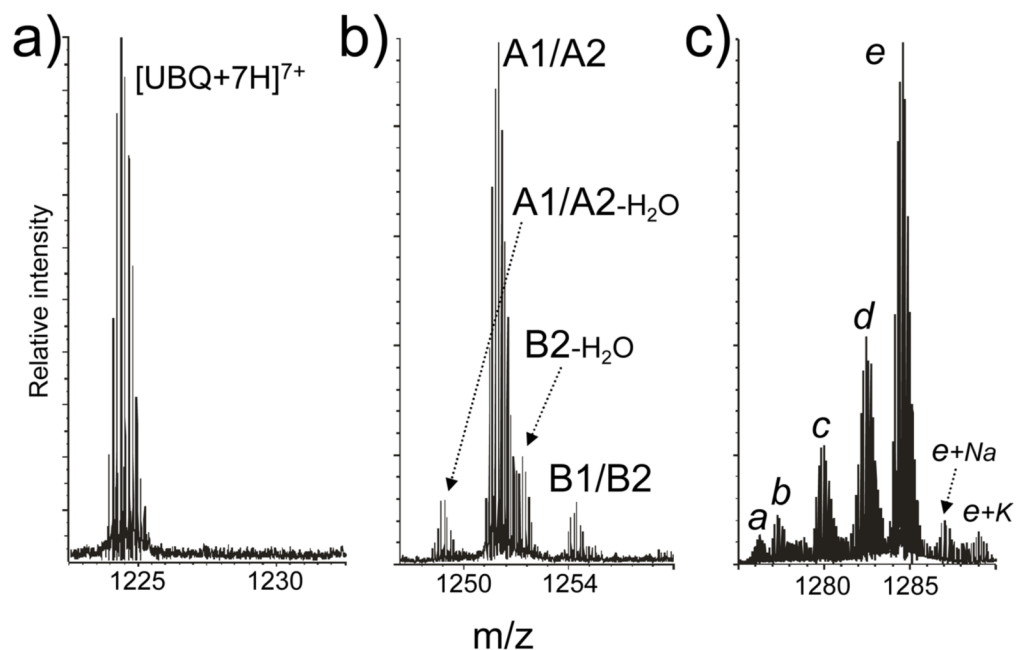


Figure 2.

Enlarged view of the regions containing unmodified ubiquitin **a**) and its mono- **b**) and bis-adducts **c**) described in Figure 1b. These data exemplify the challenge posed by the possible presence of multiple combinations of isobaric products (see the text). In panel **b**), the species detected with 8731.57 Da monoisotopic mass could match either A1-H₂O or A2-H₂O, which share a calculated mass of 8731.64 Da ($\Delta m = 172.02$ Da, Scheme 2). Similarly, species with 8749.58 and 8775.57 Da monoisotopic mass could match A1/A2 and B1/B2, which present calculated masses of 9749.65 and 8775.65 Da, respectively (Δm of 190.03 and 216.03 Da, Scheme 2). In contrast, species with experimental mass 8757.57 Da (8757.64 Da calc.) can be unambiguously assigned to B2-H₂O ($\Delta m = 198.02$ Da, Scheme 2). In panel **c**), species **a** detected with monoisotopic mass 8921.52 Da could contain two adducts consisting of A1/A2 and A1/A2-H₂O (incremental masses listed in Scheme 2); species **b** (experimental mass 8929.53 Da) could contain both **B1/B2** and **A2-2H₂O**; species **c** (exp. mass 8947.51 Da) could contain both **B1/B2** and **A1-H₂O/A2-H₂O**; species **d** (exp. mass 8965.52 Da) could contain both **B1/B2** and **A1/A2**; and species **e** (exp. mass 8983.55 Da) could contain both **B1/B2** and **A3**.

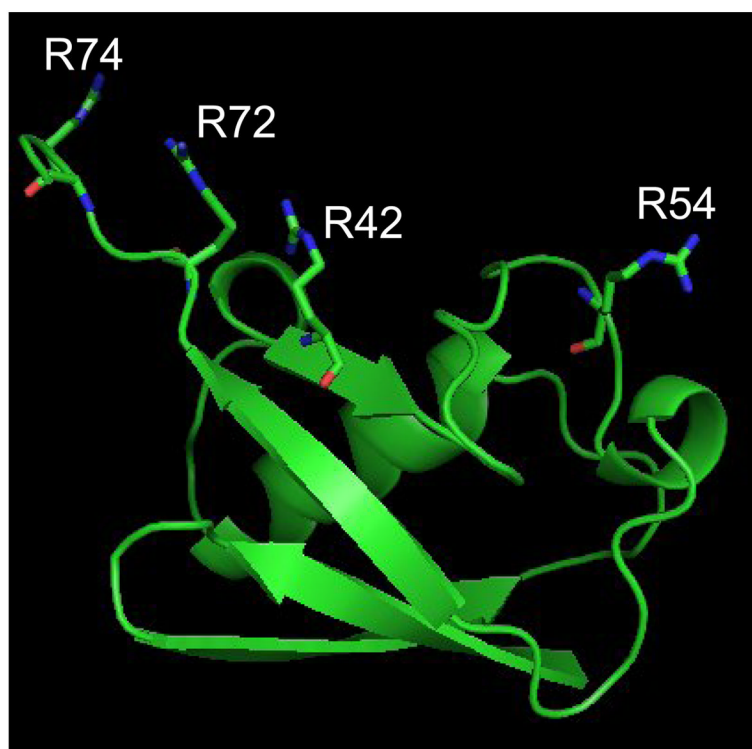


Figure 3. The high-resolution structure of bovine erythrocytes ubiquitin (PDB ID: 1ARR) [69] shows that R42, R72, and R74 are brought close together by the protein fold, whereas R54 remains isolated and out of crosslinking range.

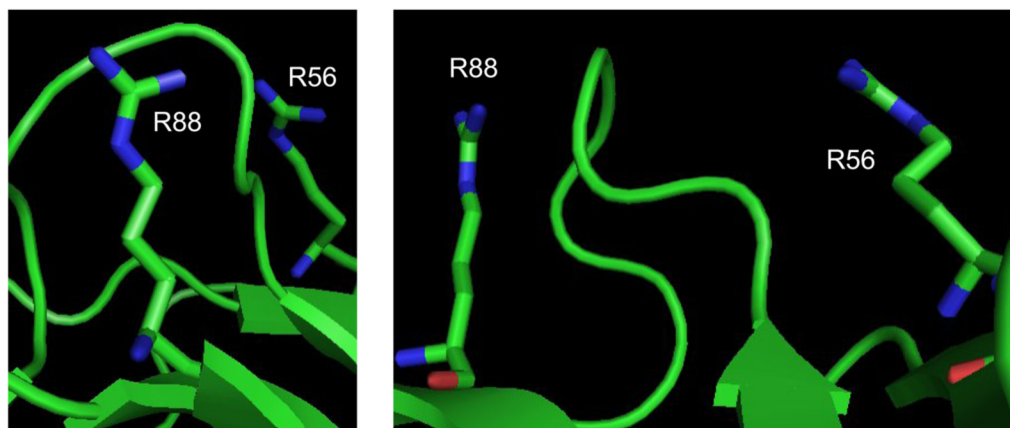


Figure 4. Front and side view of the region containing R56 and R88 of bovine erythrocytes carbonic anhydrase II (PDB ID: 1V9E)[72]. The line of sight between the target side-chains is obstructed by the loop comprised between Y69 and V77. The formation of R56-R88 crosslinks can take place only if the susceptible arginines and intervening loop are capable of adopting alternative conformations in solution.

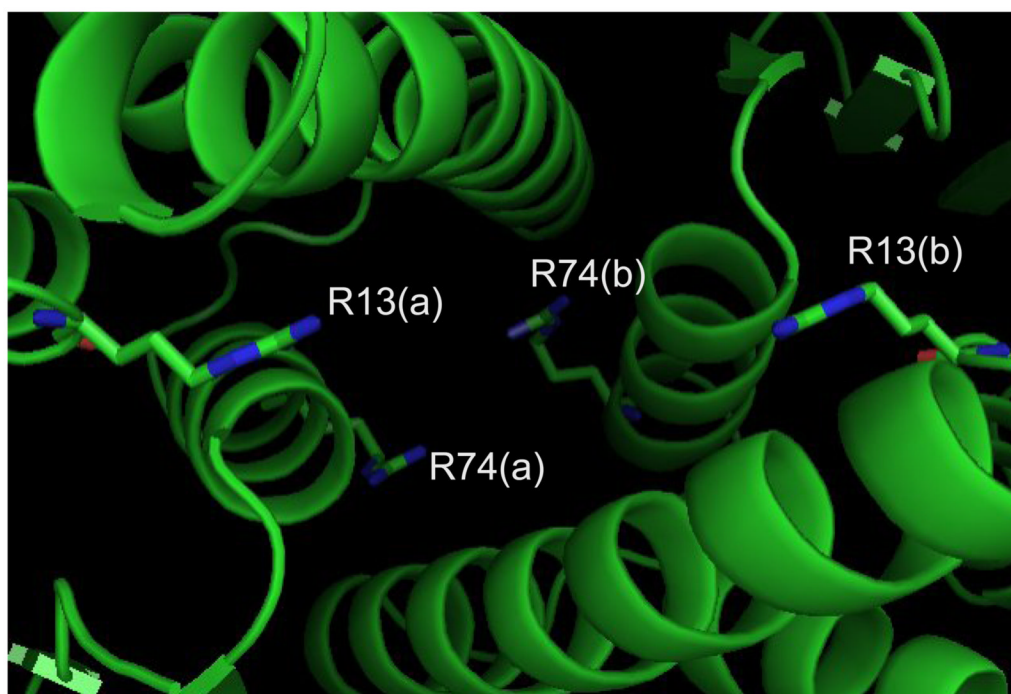
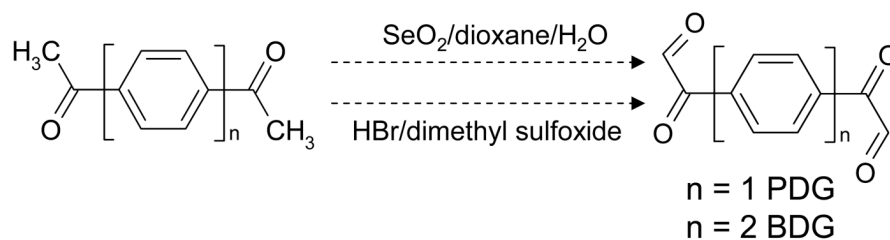
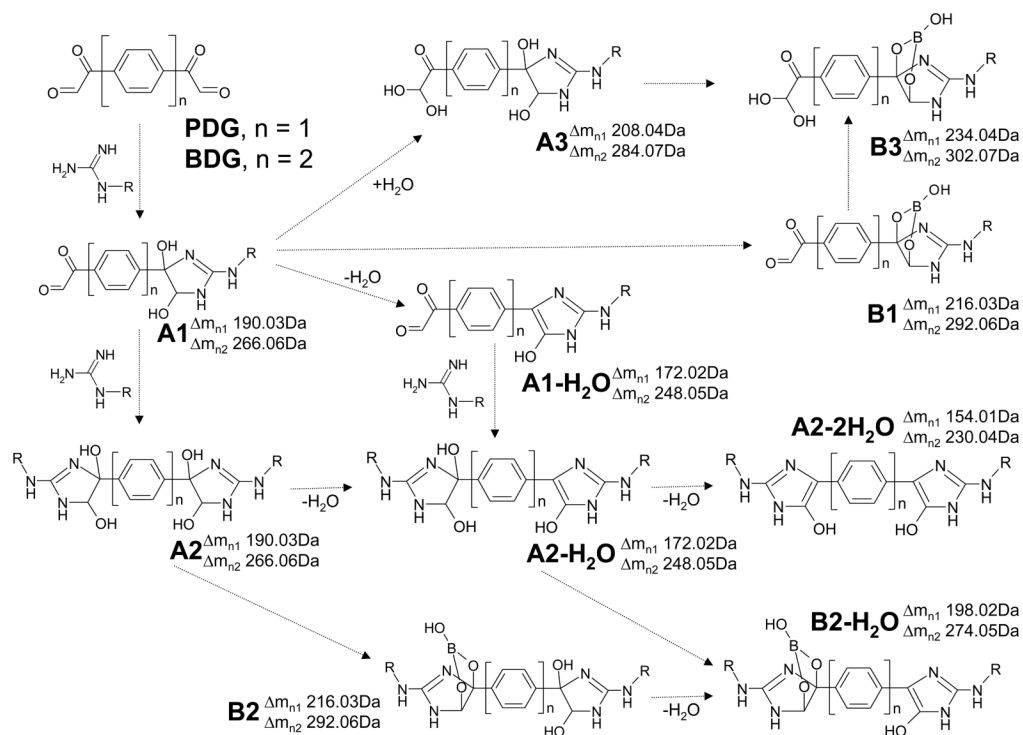


Figure 5. Detail of the interface between subunits in human placenta pi-class glutathione S-transferase (PDB ID: 14GS)[73]. In the foreground, residues R13(a) and R13(b) of cognate subunits face each other across the gap, in ideal positions for the formation of inter-molecular bridging adducts. In the background, R74(a) and R74(b) are also situated in favorable positions for inter-subunit crosslinking.

**Scheme 1.**

The bifunctional crosslinkers 4-phenyl-diglyoxal (PDG, n=1) and 4,4'-biphenyl-diglyoxal (BDG, n=2) were obtained by oxidation of 1,4-diacetylbenzene and 4,4'-diacetylbiphenyl, respectively. Initially, two alternative procedures were evaluated, which employed either selenium dioxide (SeO₂) in dioxane/water, or hydrobromic acid (HBr) in dimethyl sulfoxide. After these tests, the latter was selected for its ability to provide cleaner reagent (see text). PDG exhibits 6.14 ± 0.64 Å between reactive 1,2-diol functions, while BDG has a 10.44 ± 0.80 Å spacing.



Scheme 2.

Typical products observed in ESI-FTICR mass spectra of reaction mixtures obtained by treating arginine-containing substrates with 1,4-phenyl-diglyoxal (PDG, n = 1) and 4,4'-biphenyl-diglyoxal (BDG, n = 2). Adducts including 1,2-diol functions (*i. e.*, **A1**, **A2**, and **A3**) were observed together with their respective borate diesters (*i. e.*, **B1**, **B2** and **B3**). Dashed arrows indicate possible formation pathways that could take place either in solution or during analysis.

Table 1

Sequences of model substrates employed in the study: bovine erythrocytes ubiquitin (UBQ), bovine erythrocytes carbonic anhydrase II (CA), and human placenta pi-class glutathione S-transferase (GST). Arginine residues are highlighted as possible alkylation sites.

Protein	Statistics	Sequence
UBQ	76 residues 4 arginines 8559.62 Da	MQIFVKLTIG KTITLVEPS DTIENVKAKI QDKEGIPPDQ QRLIFAGKQL EDG R TLSDYN IQKESTLHLV LRLRGG
CA	259 residues 9 arginines 28964.67 Da	SHHWGYGKHN GPEHWHKDFP IANGER R QSPV DIDTKAVVQD PALKPLALVY GEATS RR MVN NGHSEFNVEYD DSQDKAVLKD GPLTGT YRLV QFHFHWGSSD DQGSEHTV D KKYAAELHLV HWNTKYGDFG TAAQQPDGLA VVGVFLKVG ANPALQKVL D ALDSIKTKGK STDFPNFDPG SLLPNVLDYW TYPGSLTTPP LLESVTWIVL KEPISVSSQQ MLK F TLNFN AEGEPELLML ANW R PAQPLK NRQVRG FPK
GST	209 residues 8 arginines 23209.98 Da	PPYTVVYFPV RGRCAALRML LADQGQSWKE EVVTVETWQE GSLKASCLYG QLPKFQDGD L TLYQSN TILR HLG R TLGLYG KDQQAALVD MVNDGVED L CKYISLIYTN YEAGKDDYVK ALPGQLKPF TLLSQNGGK TFIGDQISF ADYNLLDLLL IHEVLAPGCL DAFPLLSAYV GRLSAR PKLK AFLASPEYVN LPINGNGKQ

Table 2

Summary of mono- (Mo) and bifunctional (Bi) adducts obtained by treating bovine ubiquitin (UBQ) with either PDG or BDG, followed by digestion with trypsin (T) or chymotrypsin (C). Monoisotopic masses were calculated from the incremental masses described in Scheme 2 and from theoretical peptide masses provided by the PeptideMass calculator [67]. Note that the species with mass 3600.8018 Da corresponds to products in which either R42 or R54 were mono-functionally modified ($\Delta M = 172.02$ Da). The formation of a bifunctional conjugate bridging these residues was ruled out by the absence of diagnostic products with $\Delta M = 154.01$ or 198.02 Da.

Reagent	Obs. mass (Da)	Calc. mass (Da)	Digestion product	Modified site(s)
PDG	923.4965	923.4972	70–76 + Bi, C	R72 ↔ R74
	1603.8461	1603.8466	64–76 + Bi, T	R72 ↔ R74
	3600.8018	3600.7847	34–63 + Mo, T	R42 or R54
	1983.9332	1983.9213	44–59 + Mo, C	R54
	991.5003	991.5072	68–73 + Mo, C	R72
BDG	1551.7540	1551.7511	68–76 + 2Mo, C	R72 & R74
	2026.9344	2026.9269	49–63 + Mo, T	R54
	3175.5915	3175.5822	46–71 + Mo, C	R54
	1015.5450	1015.5487	68–73 + Mo, C	R72
	536.2050	536.2016	74–76 + Mo, C	R74

Table 3

Summary of mono- (Mo) and bifunctional (Bi) adducts obtained by treating bovine carbonic anhydrase II (CA) with either PDG or BDG, followed by digestion with trypsin (T) or chymotrypsin (C). Monoisotopic masses were calculated from the incremental masses described in Scheme 2 and from theoretical peptide masses provided by the PeptideMass calculator [67]. Note that the species with mass 2392.0974 Da corresponds to products in which either R56 or R57 were mono-functionally modified ($\Delta M = 284.07$ Da). The formation of a bifunctional conjugate bridging these residues was ruled out by the absence of diagnostic products with $\Delta M = 230.04$ or 274.05 Da.

Reagent	Obs. mass (Da)	Calc. mass (Da)	Digestion product	Modified site(s)
PDG	5995.9628	5995.9518	36–88 + Bi, T	R56 ↔ R57
	2391.0906	2391.1138	51–58/88–96 + Bi, C	R56 (R57) ↔ R88
	3815.7660	3815.7484	57–88 + Mo, T	R57
	4592.1720	4592.1684	89–125 + Mo, T	R110
	2985.4593	2985.4615	18–35/253–259 + Bi, T	R26 ↔ R255
BDG	2392.0974	2392.0911	47–65 + Mo, C	R56 or R57
	4271.0332	4271.0166	49–69/79–92 + Bi, C	R56 (R57) ↔ R88
	4612.1528	4612.1509	49–78/88–94 + Bi, C	R56 (R57) ↔ R88
	3105.4148	3105.4001	89–112 + Mo, T	R110

Table 4

Summary of mono- (Mo) and bifunctional (Bi) adducts obtained by treating dimeric human glutathione S-transferase (GST) with either PDG or BDG, followed by digestion with trypsin (T) or chymotrypsin (C). Indexes (a) and (b) indicate cognate subunits. Monoisotopic masses were calculated from the incremental masses described in Scheme 2 and from theoretical peptide masses provided by the PeptideMass calculator [67].

Reagent	Obs. mass (Da)	Calc. mass (Da)	Digestion product	Modified site(s)
PDG	2809.4776	2809.4880	73–78(a)/63–79(b) + Bi, C	R74(a) ↔ R74(b)
	860.4357	860.4396	184–189 + Mo, C	R186
BDG	2809.4776	2809.4819	9–17(a)/8–21(b) + Bi, C	R13(a) ↔ R13(b)
	2300.2505	2300.2400	73–78/9–21 + Bi, C	R74 ↔ R13 or R74 ↔ R11
	1646.8526	1646.8571	177–189 + Bi, C	R182 ↔ R186
	1815.9114	1815.8984	1–13 + Mo, T	R11
	1778.8734	1778.8649	18–19/63–72 + Bi, C	R18 ↔ R70