



computational proteomics

## Laboratory for Computational Proteomics

[www.FenyoLab.org](http://www.FenyoLab.org)

E-mail: [Info@FenyoLab.org](mailto:Info@FenyoLab.org)

Facebook: [NYUMC Computational Proteomics Laboratory](#)

Twitter: [@CompProteomics](#)

## A Modular Cross-Linking Approach for Exploring Protein Interactions

Michelle Trester-Zedlitz,<sup>†,‡</sup> Katsuhiko Kamada,<sup>§</sup> Stephen K. Burley,<sup>||</sup> David Fenyö,<sup>⊥</sup>  
Brian T. Chait,<sup>\*,‡</sup> and Tom W. Muir<sup>\*,†</sup>

*Contribution from the Laboratories of Synthetic Protein Chemistry, Mass Spectrometry and Gaseous Ion Chemistry, and Molecular Biophysics, Rockefeller University, 1230 York Avenue, New York, New York 10021*

Received May 15, 2002; E-mail: muilt@mail.rockefeller.edu, chait@mail.rockefeller.edu

**Abstract:** A method is described for the elucidation of protein–protein interactions using novel cross-linking reagents and mass spectrometry. The method incorporates (1) a modular solid-phase synthetic strategy for generating the cross-linking reagents, (2) enrichment and digestion of cross-linked proteins using microconcentrators, (3) mass spectrometric analysis of cross-linked peptides, and (4) comprehensive computational analysis of the cross-linking data. This integrated approach has been applied to the study of cross-linking between the components of the heterodimeric protein complex negative cofactor 2.

### Introduction

The elucidation of protein interactions is fundamental to understanding biological processes. Within a cell, a single protein may interact with many other molecules, and these interactions may be under both spatial and temporal regulation. Therefore, cataloguing the interactions of even a single protein represents an extremely difficult task. This challenge is further complicated by the fact that protein interactions are often quite transient in nature or may involve large multiprotein complexes. Despite these limitations, large-scale “proteomic” efforts are now under way that aim to provide a complete description of all protein interaction networks within a given cell.<sup>1,2</sup>

One approach to discovering the identity of interacting proteins is to use chemical cross-linking. Unlike other protein inter-

action methodologies,<sup>3–5</sup> chemical cross-linking irreversibly captures binding partners, so that even transient interactions can, in principle, be detected. In addition, the covalent nature of a chemical cross-link enables analytical techniques to be employed that, under normal circumstances, would disrupt native protein interactions. Two types of information can be gleaned from cross-linking studies: the identity of the interacting proteins, and the regions within those proteins that are involved in the interaction. The quality of information obtained is dependent on the accuracy, resolution, and dynamic range of the analytical strategy used to interrogate the cross-linked system. Traditionally, these readout strategies have relied heavily on chemical or enzymatic digestion of a cross-linked complex followed by gel-mobility or chromatographic separations in combination with radioactive or bioaffinity “tracers”.<sup>6,7</sup> Fragments of interest could then be analyzed further by techniques such as Edman sequencing.

Modern biological mass spectrometry (MS) provides a powerful addition to these classical readout approaches.<sup>8</sup> MS is now the method of choice for protein identification and is seeing increasing application in the area of proteomics, for example, by allowing the composition of multiprotein cellular machines to be defined.<sup>9</sup> The combination of chemical cross-linking and mass spectrometry complements these existing approaches by providing information regarding the architecture

\* To whom correspondence should be addressed. Phone: (212) 327-7368 (T.W.M.), (212) 327-8849 (B.T.C.). Fax: (212) 327-7358.

<sup>†</sup> Laboratory of Synthetic Protein Chemistry.

<sup>‡</sup> Laboratory of Mass Spectrometry and Gaseous Ion Chemistry.

<sup>§</sup> Laboratory of Molecular Biophysics.

<sup>||</sup> Current address: Structural GenomiX, Inc., 10505 Roselle St., San Diego, CA 92121.

<sup>⊥</sup> Current address: Amersham Biosciences, Corp. 800 Centennial Ave., Piscataway, NJ 08855.

- (1) Gavin, A. C.; Bosche, M.; Krause, R.; Grandi, P.; Marzioch, M.; Bauer, A.; Schultz, J.; Rick, J. M.; Michon, A. M.; Cruciat, C. M.; Remor, M.; Hofert, C.; Schelder, M.; Brajenovic, M.; Ruffner, H.; Merino, A.; Klein, K.; Hudak, M.; Dickson, D.; Rudi, T.; Gnau, V.; Bauch, A.; Bastuck, S.; Huhse, B.; Leutwein, C.; Heurtier, M. A.; Copley, R. R.; Edelmann, A.; Querfurth, E.; Rybin, V.; Drewes, G.; Raida, M.; Bouwmeester, T.; Bork, P.; Seraphin, B.; Kuster, B.; Neubauer, G.; Superti-Furga, G. *Nature* **2002**, *415*, 141–147.
- (2) Ho, Y.; Grubler, A.; Heilbut, A.; Bader, G. D.; Moore, L.; Adams, S. L.; Millar, A.; Taylor, P.; Bennett, K.; Boutillier, K.; Yang, L.; Wolting, C.; Donaldson, I.; Schandorff, S.; Shewnarane, J.; Vo, M.; Taggart, J.; Goudreault, M.; Muskat, B.; Alfarano, C.; Dewar, D.; Lin, Z.; Michalickova, K.; Willems, A. R.; Sassi, H.; Nielsen, P. A.; Rasmussen, K. J.; Andersen, J. R.; Johansen, L. E.; Hansen, L. H.; Jespersen, H.; Podtelejnikov, A.; Nielsen, E.; Crawford, J.; Poulsen, V.; Sorensen, B. D.; Matthiesen, J.; Hendrickson, R. C.; Gleeson, F.; Pawson, T.; Moran, M. F.; Durocher, D.; Mann, M.; Hogue, C. W.; Figeys, D.; Tyers, M. *Nature* **2002**, *415*, 180–183.

- (3) Miyawaki, A.; Llopis, J.; Heim, R.; McCaffery, J. M.; Adams, J. A.; Ikura, M.; Tsien, R. Y. *Nature* **1997**, *388*, 882–887.
- (4) Fields, S.; Song, O. *Nature* **1989**, *340*, 245–246.
- (5) Anderson, N. G. *Methods Mol. Biol.* **1998**, *88*, 35–45.
- (6) Bickle, T. A.; Hershey, J. W.; Traut, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 1327–1331.
- (7) Wold, F. *Methods Enzymol.* **1972**, *25*, 623–651.
- (8) Bennett, K. L.; Kussmann, M.; Bjork, P.; Godzwon, M.; Mikkelsen, M.; Sorensen, P.; Roepstorff, P. *Protein Sci.* **2000**, *9*, 1503–1518 and references therein.
- (9) Mann, M.; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437–473.

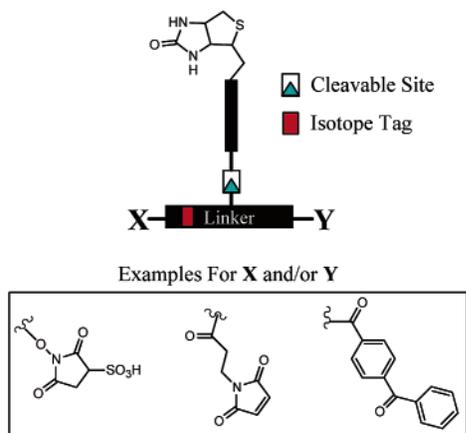


Figure 1. Modular cross-linking reagent design.

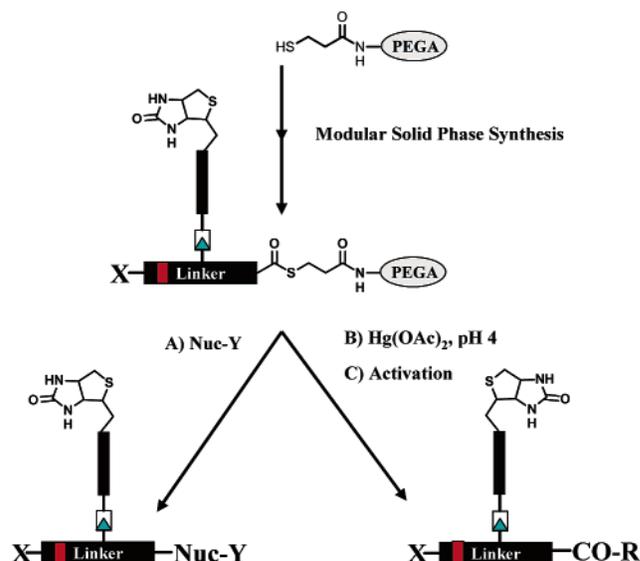
of large multicomponent complexes.<sup>10</sup> Elucidation of higher resolution structural information, for example, defining specific binding surfaces, may also be possible in some cases.<sup>11</sup>

Mass analyses of digested cross-linked proteins are complicated by sample heterogeneity; in most cases samples contain a complex mixture of modified and unmodified peptides.<sup>8</sup> Potential solutions to this problem include procedures that either enrich the cross-linked products or provide them with a recognizable mass signature. Several research groups have found that the inclusion of chemical moieties, such as affinity handles,<sup>12</sup> isotope tags,<sup>13,14</sup> and cleavable sites,<sup>8</sup> within their modifying or cross-linking reagent can facilitate MS analyses. At present, very few cross-linking reagents with all the aforementioned properties are commercially available.<sup>15</sup>

We are interested in the development of chemical cross-linking strategies that employ MS as the readout step. Ideally, one would like to design a general strategy and a single reagent that can be used for all cross-linking studies. However, this is probably unrealistic due to the structural diversity among proteins; no two proteins will have exactly the same three-dimensional distribution of reactive side chains. Thus, it is unlikely that a single reagent will work for every application. Moreover, it may be useful to probe any given system with a variety of reagents each with different properties. To address these needs, we have developed a strategy, the key feature of which is a rapid, modular approach to the synthesis of cross-linking reagents. Modularity allows the properties of the reagent to be systematically varied (Figure 1), making it feasible to optimize the reagent for each different cross-linking application.

In this paper, we describe the syntheses of five trifunctional cross-linkers, all of which were designed to be flexible and soluble in water and to contain a biotin affinity handle. To target different protein functional groups and enhance MS analyses, several properties were varied within the cross-linker, specifically, spacer lengths, chemical reactivities, cleavage sites, and

Scheme 1. Use of a Multidetachable Thioester Linker



isotope tags. Although we applied several of our reagents to the heterodimeric protein complex negative cofactor 2 (NC2) (the structure of which is known),<sup>16</sup> we only provide details on the most efficient of these, i.e., **4** and **6**. The studies of NC2 employed a novel “one-pot” procedure that allowed straightforward isolation of cross-linked protein fragments suitable for MS analysis. Interpretation of the MS data was facilitated by a specialized computer program which exhaustively compares experimental masses to all possible<sup>17</sup> calculated cross-linked combinations. Importantly, in this paper we highlight the utility of mass spectrometry as a continuous feedback tool for optimizing cross-linking syntheses and activation, and for following the course of the cross-linking experiment.

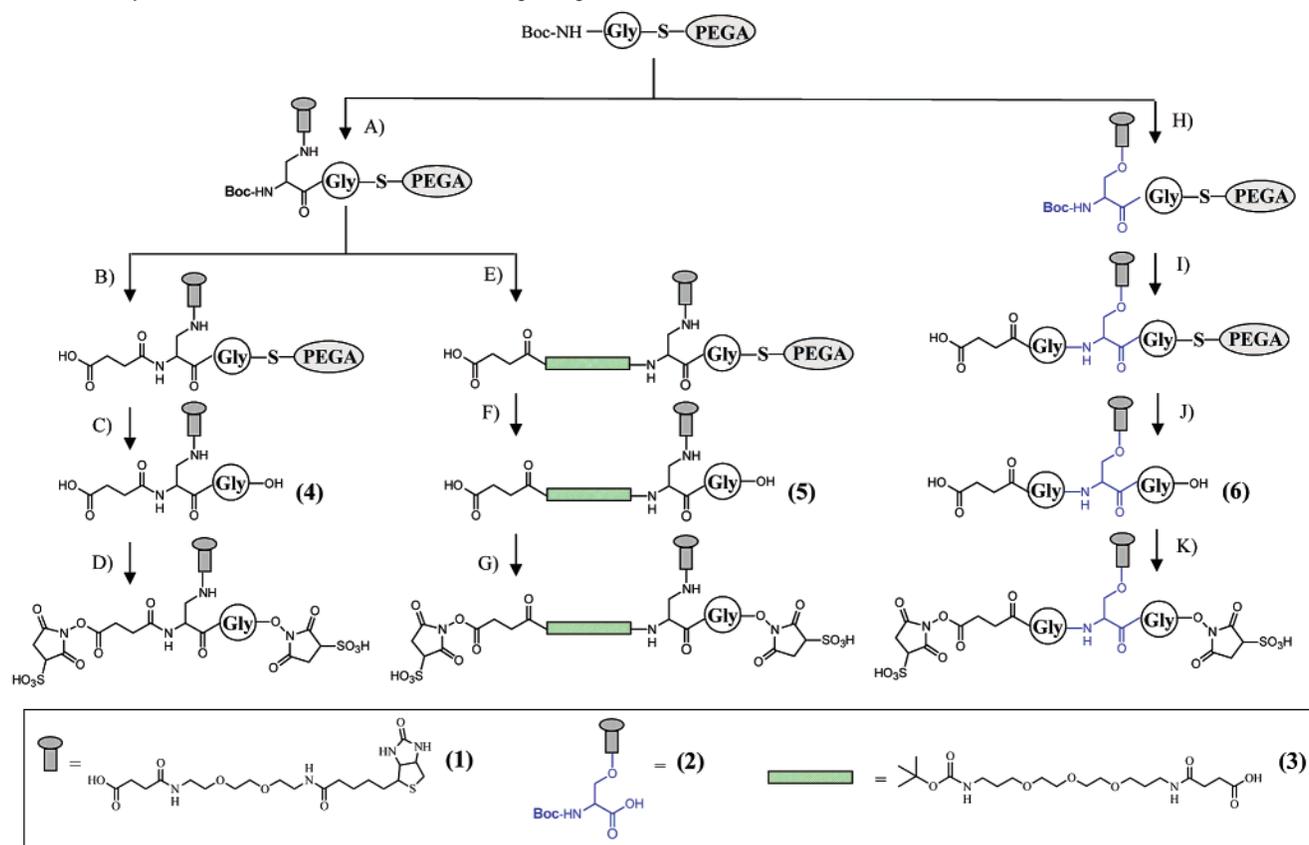
## Results

**Cross-Linker Syntheses.** Cross-linking reagents were synthesized using solid-phase chemistry.<sup>18</sup> The synthetic strategy was tailored to be compatible with isotope labels, affinity handles, and a variety of reactive groups at the extremities of the reagent. Moreover, the syntheses were designed such that the end product would be obtained directly from the resin upon cleavage with nucleophiles (i.e., no global deprotection steps were required).

The solid-phase approach outlined in Scheme 1 exploits the chemical versatility of the mercaptopropionamide–PEGA resin linker.<sup>19</sup> An important feature of this system is the acid stability of the alkyl thioester linkage formed between the resin linker and the cross-linker. Reagents can thus be prepared on the solid phase using standard *tert*-butyl protection schemes. Use of the PEGA resin is also important to our strategy since it is compatible with both organic and aqueous solvent systems.<sup>20</sup>

- (10) Rappsilber, J.; Siniosoglou, S.; Hurt, E. C.; Mann, M. *Anal. Chem.* **2000**, *72*, 267–275.  
 (11) Young, M. M.; Tang, N.; Hempel, J. C.; Oshiro, C. M.; Taylor, E. W.; Kuntz, I. D.; Gibson, B. W.; Dollinger, G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 5802–5806.  
 (12) Alley, S. C.; Ishmael, F. T.; Jones, A. D.; Benkovic, S. J. *J. Am. Chem. Soc.* **2000**, *122*, 6126–6127.  
 (13) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994–999.  
 (14) Mueller, D. R.; Schindler, P.; Towbin, H.; Wirth, U.; Voshol, H.; Hoving, S.; Steinmetz, M. O. *Anal. Chem.* **2001**, *73*, 1927–1934.  
 (15) Hermanson, G. T. *Bioconjugate Techniques*; Academic Press: New York, 1995.

- (16) Kamada, K.; Shu, F.; Chen, H.; Malik, S.; Stelzer, G.; Roeder, R. G.; Meisterernst, M.; Burley, S. K. *Cell* **2001**, *106*, 71–81.  
 (17) On the basis of the extent of cross-linking and enzymatic digestion observed with MS, the user can limit the search in the following ways: (1) maximum number of fragments cross-linked; (2) maximum number of cross-linker attachments; (3) maximum number of missed enzymatic cleavage sites; (4) maximum number of partial global modifications (e.g., Cys modifications).  
 (18) Merrifield, R. B. *J. Am. Chem. Soc.* **1963**, *85*, 2149.  
 (19) Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *J. Pept. Res.* **1998**, *51*, 303–316.  
 (20) Meldal, M. *Tetrahedron Lett.* **1992**, *33*, 3077–3080.

**Scheme 2.** Synthesis of Amine-Reactive Cross-Linking Reagents<sup>a</sup>

<sup>a</sup> Conditions: (A) (i) TFA; (ii) Boc-Dapa(Fmoc), HBTU, DIEA, DMF; (iii) 1% DBU; (iv) **1**, HBTU, DIEA, DMF; (B) (i) TFA; (ii) succinic anhydride, DIEA, DMF; (C) 1 M NaOH; (D) (i) tributylamine; (ii) sulfo-NHS, PyBOP, DIEA, DMF; (E) (i) TFA; (ii) **3**, HBTU, DIEA, DMF; (iii) TFA; (iv) succinic anhydride, DIEA, DMF; (F) 1 M NaOH; (G) (i) tributylamine; (ii) sulfo-NHS, PyBOP, DIEA, DMF; (H) (i) TFA; (ii) **2**, PyBOP, DIEA, DMF; (I) (i) TFA; (ii) Boc-Gly-OH, HBTU, DIEA, DMF; (iii) TFA; (iv) succinic anhydride, DIEA, DMF; (J) 30 mM Hg(OAc)<sub>2</sub>, 30% MeCN, 5% glacial acetic acid; (K) (i) CsCl; (ii) sulfo-NHS, PyBOP, DIEA, DMF.

This greatly expands the range of cleavage conditions that can be used to remove the final product from the solid support (Scheme 1). For example, mild aqueous hydrolysis conditions (Hg(OAc)<sub>2</sub>, pH 4)<sup>21</sup> can be used, thereby allowing the inclusion of base-labile components in the cross-linking reagents.

To demonstrate the versatility of the mercaptopropionamide-PEGA system, five different cross-linking reagents were designed and synthesized. The amine-reactive trifunctional cross-linking reagents' (**4–6**) syntheses are shown in Scheme 2. All three syntheses began with the attachment of Boc-Gly to the solid support through a thioester linkage. For reagents **4** and **5**, Boc-Dapa(Fmoc) was added using standard in situ activation procedures.<sup>22</sup> Selective removal of the Fmoc group from the  $\beta$ -amino group of the Dapa moiety allowed attachment of biotin derivative **1**. Deprotection of the base-labile Fmoc protecting group, without unwanted aminolysis of the thioester linkage, was accomplished using a modified version of the DBU-based procedure described by Wade and co-workers.<sup>23</sup> The spacer length in reagent **5** was increased by the attachment of trioxa compound **3** to the Dapa  $\alpha$ -amino group. Following N $^{\alpha}$  deprotection and acylation with succinic anhydride, compounds **4** and **5** were obtained by treatment of the resin with aqueous

NaOH. The solid-phase synthesis of cross-linker **6** employed Boc-serine derivative **2** in which the  $\beta$ -OH group is esterified with **1**. The presence of the ester group prevented the use of NaOH in the cleavage step. However, hydrolysis using Hg(OAc)<sub>2</sub>, pH 4, was found to proceed with excellent efficiency (89%)<sup>24</sup> and without detectable cleavage of the ester.

The synthesis of the isotope-tagged trifunctional cross-linking reagent **8a,b** is shown in Scheme 3. This reagent was made in a stepwise fashion similar to that of **4** with the following exceptions: 4-benzoylbenzoic acid was attached to the  $\beta$ -amino group of Dapa, and the N $^{\alpha}$  group of Dapa was acylated with either succinic anhydride, **8a**, or 2,2',3,3'-*d*-succinic anhydride, **8b**. Nucleophilic cleavage of the reagents from the solid support was achieved by treatment with biotin derivative **7**. The aminolysis step was quantitative as indicated by reversed-phase HPLC analysis of the crude material.<sup>24</sup>

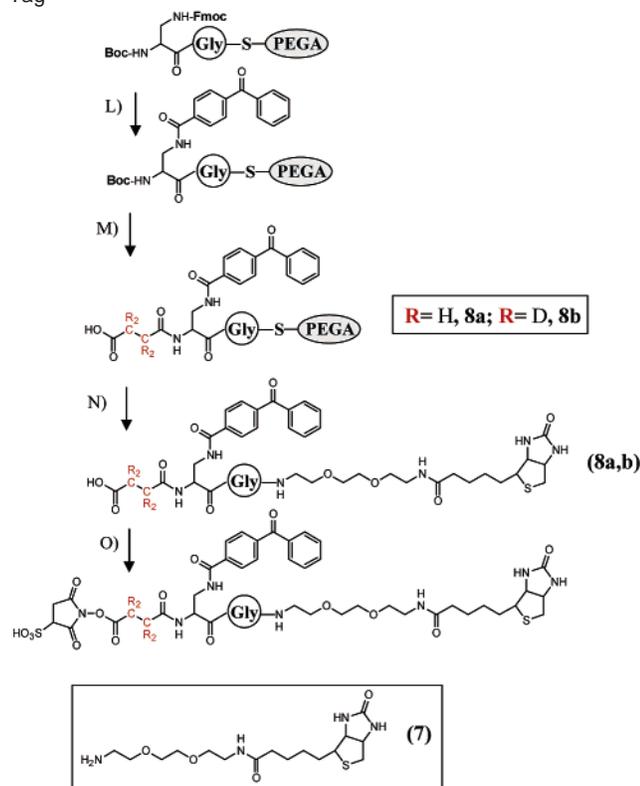
Trifunctional reagent **9** was synthesized according to Scheme 4. This reagent was made similarly to reagent **4** except for the replacement of succinic anhydride with 3-maleimidopropionic acid. Due to the lability of maleimides at high pH, reagent **9** was released from the support with Hg(OAc)<sub>2</sub> at pH 4 with a cleavage efficiency of 92%.<sup>24</sup> Importantly, the maleimide group was found to be stable to these cleavage conditions as indicated by HPLC and MS.

(21) Camarero, J. A.; Ayers, B.; Muir, T. W. *Biochemistry* **1998**, *37*, 7487–7495.

(22) Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. *Int. J. Pept. Protein Res.* **1992**, *40*, 180–193.

(23) Wade, J. D.; Bedford, J.; Sheppard, R. C.; Tregear, G. W. *Pept. Res.* **1991**, *4*, 194–199.

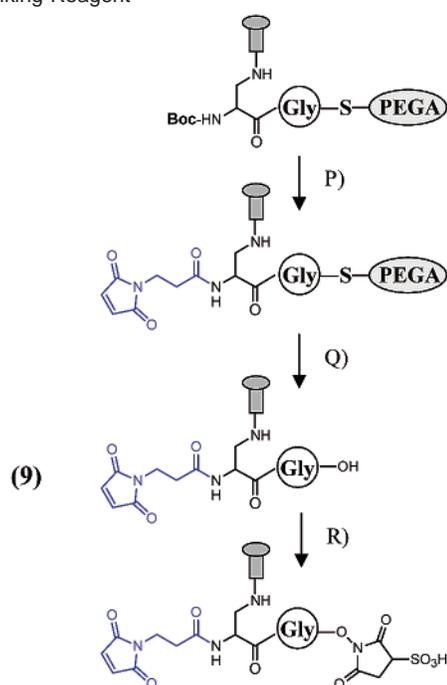
(24) Cleavage supernatants were analyzed by reversed-phase HPLC, and relevant peak areas were compared with those obtained by subsequent NaOH cleavage of the same beads.

**Scheme 3.** Synthesis of a Cross-Linking Reagent with an Isotope Tag<sup>a</sup>

<sup>a</sup> Conditions: (L) (i) 1% DBU, DMF; (ii) 10% DIEA; (iii) 4-benzoylbenzoic acid, DIC, DMF; (M) (i) TFA; (ii) succinic anhydride or 2,2',3,3'-d<sub>4</sub>-succinic anhydride, DIEA, DMF; (N) **7**, DMF, DIEA; (O) (i) tributylamine, (ii) sulfo-NHS, PyBOP, DIEA, DMF.

**Cross-Linker Activation.** In preliminary studies, we explored a variety of conditions for activation of carboxyl groups within our cross-linking reagents. These strategies included the use of DIC with NHS, EDC with sulfo-NHS, and PyBOP with both NHS and sulfo-NHS. The efficiency of these activation procedures was followed using an arginine trapping procedure which allowed for straightforward MS analysis of the acylated products. None of these activation procedures gave high yields of the desired products when the free acid form of the cross-linking reagents was used. A typical result is shown in Figure 2A in which PyBOP and sulfo-NHS were used in the activation of reagent **4**. MS analysis indicated that the major product contained a single arginine addition, and tandem MS analysis further revealed that this acylation had selectively taken place on the glycine carboxyl group (Figure 2C,E). As shown in Figure 2B, the yield of the diacylated adduct was greatly improved when a tributylammonium or cesium salt of the diacid was employed in the activation step. Presumably, salt formation improved the nucleophilicity of the succinic acid carboxyl group relative to the glycyl carboxyl group.<sup>25</sup>

**Cross-Linking of the NC2 Heterodimer.** NC2 is a protein heterodimer (consisting of  $\alpha$  and  $\beta$  subunits) that functions as a negative regulator of basal transcription by blocking the assembly of the preinitiation complex. The X-ray structure of an NC2 construct, incorporating residues 1–77 of the  $\alpha$  subunit and 1–176 of the  $\beta$  subunit, in complex with TATA box binding

**Scheme 4.** Synthesis of a Sulfhydryl- and Amine-Reactive Cross-Linking Reagent<sup>a</sup>

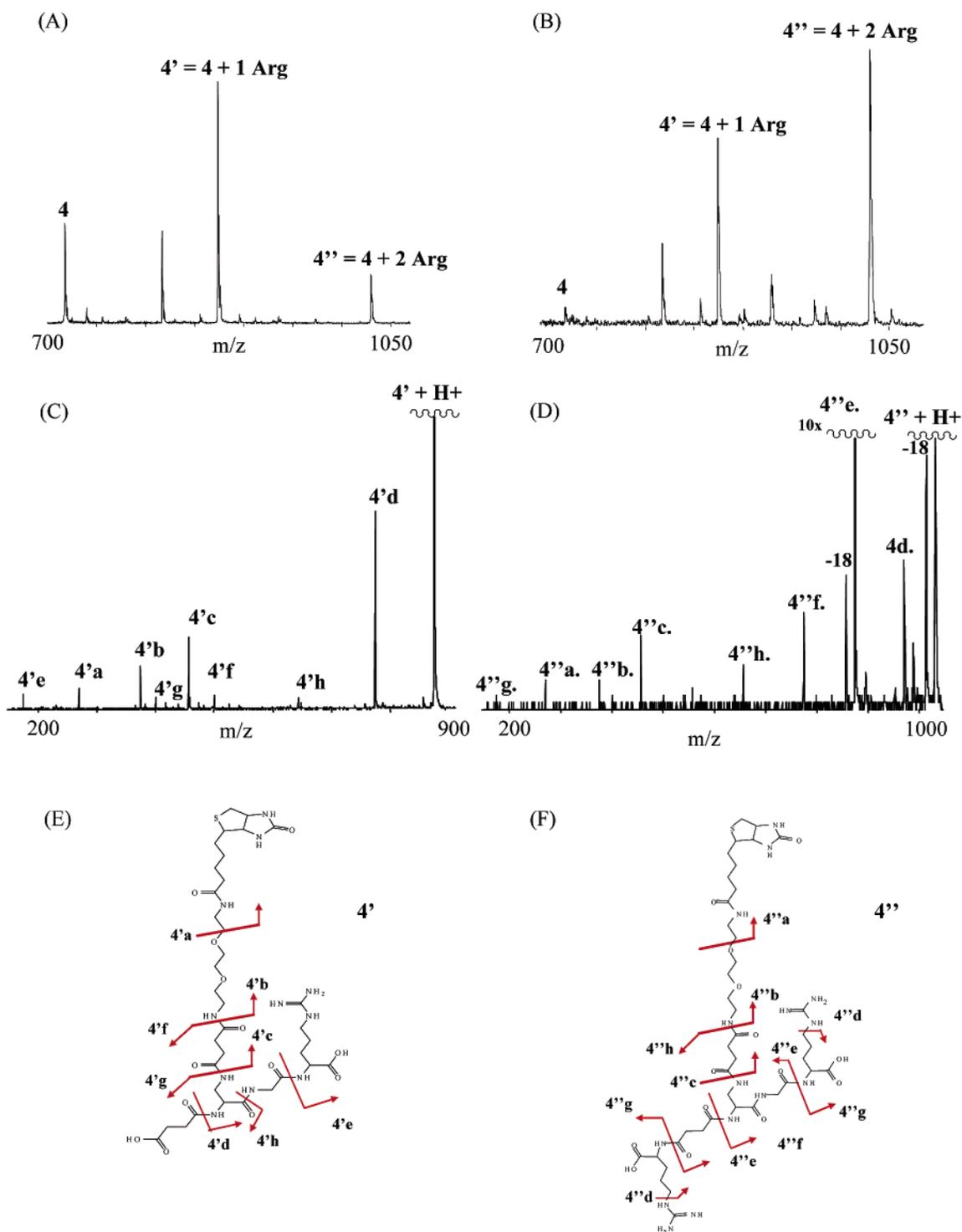
<sup>a</sup> Conditions: (P) (i) TFA; (ii) 3-maleimidopropionic acid, PyBOP, DIEA, DMF; (Q) 30 mM Hg(OAc)<sub>2</sub>, 30% MeCN, 5% glacial acetic acid; (R) sulfo-NHS, PyBOP, DIEA, DMF.

protein (TBP) and promoter DNA has recently been determined at 2.6 Å resolution.<sup>16</sup> As a model system for our cross-linking studies, we chose a related NC2 complex, incorporating residues 11–95 of the  $\alpha$  subunit and 7–99 of the  $\beta$  subunit, which was used in early efforts to crystallize the complex. Although crystals of this latter form were not obtained, biochemical analyses indicated that the two polypeptides form a tight complex in solution (data not shown), which we anticipated would be closely related to the corresponding NC2 component of the NC2–TBP–DNA crystal structure.<sup>16</sup>

The truncated NC2  $\alpha\beta$  heterodimer was combined with a ~70-fold molar excess of activated **4**, **6**, and **8a** on ice at pH 7 (Figure 3A (i)). These conditions were empirically selected on the basis of a desire to attain single-hit conditions with the highest yield of cross-linked heterodimer. SDS–PAGE and mass spectrometry results of **8a** with NC2 indicated that this cross-linker was less efficient and yielded reaction products that were significantly more difficult to analyze than that with **4** or **6**. Therefore, we will focus only on results using the latter reagents.

Aliquots were taken at four different time points for analyses by SDS–PAGE (Figure 3B) and MALDI-TOF MS (Figure 3C). Each time point was immediately quenched with 2-mercaptoethanol and then acidified. The SDS–PAGE analysis in Figure 3B provides a quantitative assessment of the NC2 cross-linking time course. At just 5 min, a gel shift, composed of one  $\alpha\beta$  complex covalently tethered with reagent **4** or **6**, was observed. Estimated cross-linking efficiencies of the time course (5–60 min) experiment ranged from 10% to 25%. MS analysis in Figure 3C provides a qualitative assessment of the extent of cross-linking within the  $\alpha\beta$  complex. The mass spectra show that, by 5 min, a portion of the  $\alpha$  and  $\beta$  monomers have become

(25) Dijkstra, G.; Kruijzinga, W. H.; Kellogg, R. M. *J. Org. Chem.* **1987**, *52*, 4230–4234.



**Figure 2.** Trapping the activated cross-linking reagent with arginine: (A) MALDI-TOF MS of Arg-quenched sulfo-NHS-activated **4**; (B) MALDI-TOF MS of the Arg-quenched sulfo-NHS-activated tributylamine salt of **4**; (C) MALDI-QqTOF MS/MS of  $4' = [4 + \text{Arg}]$ ; (D) MALDI-QqTOF MS/MS of  $4'' = [4 + 2 \text{Arg}]$ ; (E) proposed cleavages inferred from MS/MS data of  $4' = [4 + \text{Arg}]$ ; (F) proposed cleavages inferred from MS/MS data of  $4'' = [4 + 2 \text{Arg}]$ .

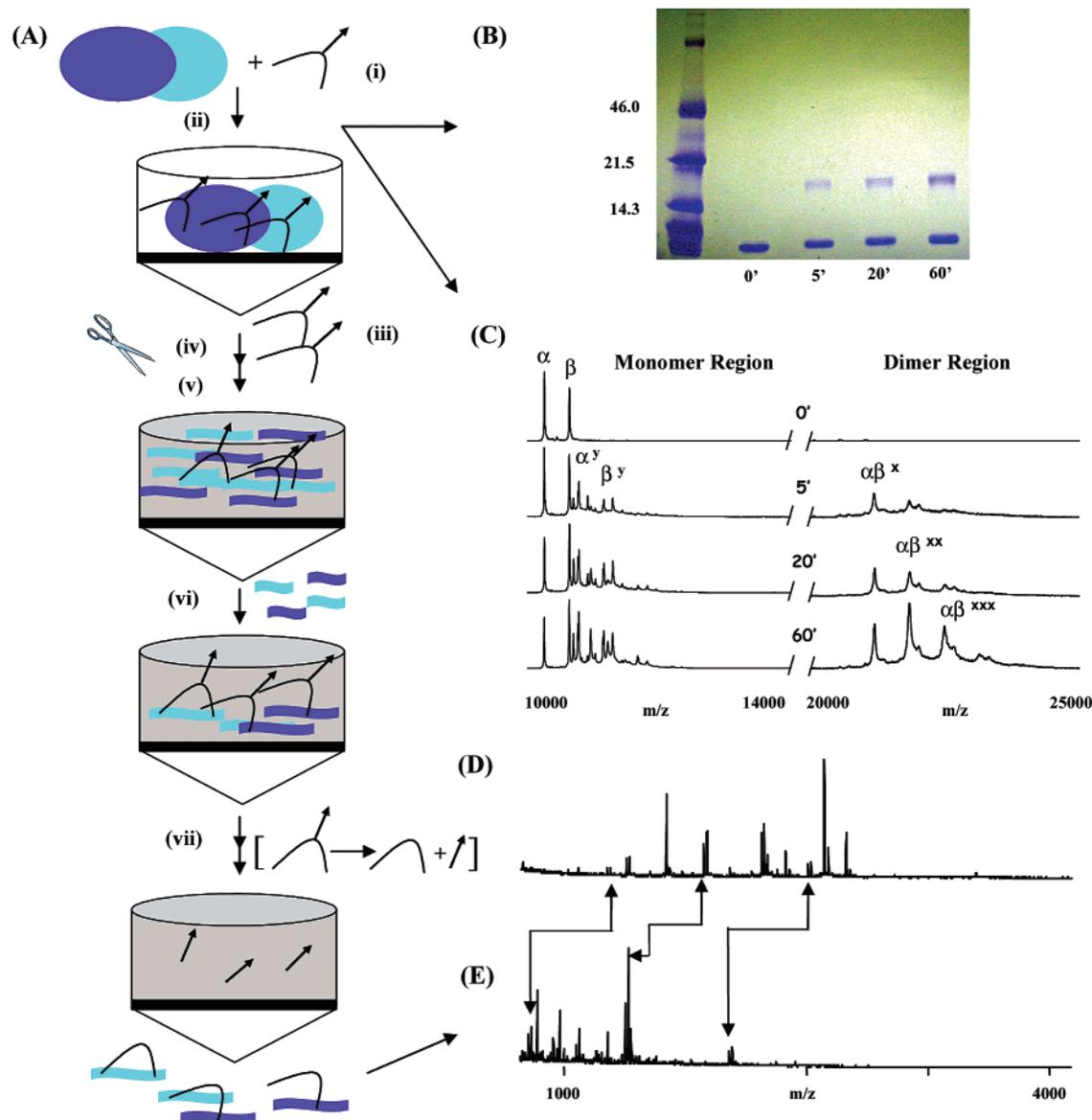
tethered together with one cross-link. And by 60 min, up to three cross-links were evident.<sup>26</sup>

**Analysis of Cross-Linked Proteins.** To isolate cross-linked protein fragments suitable for MS analysis, we devised the procedure that is summarized in Figure 3A. Importantly, microconcentrators were used as a porous surface to facilitate

excess cross-linker removal, retention of cross-linked protein complexes, easy exchange of experimental conditions, and the potential for size exclusion.<sup>27</sup> MS provided a quick and informative tool for assessing each step of the process. Upon searching for combinations of cross-linked protein digest fragments, it became apparent that we were looking for needles in a haystack. To locate and identify these “needles”, we utilized

(26) Observation of the cross-linked product requires that only one of the three be intermolecular. The other two links can be either inter- or intramolecular.

(27) Note: Some sample loss on filters should be anticipated.

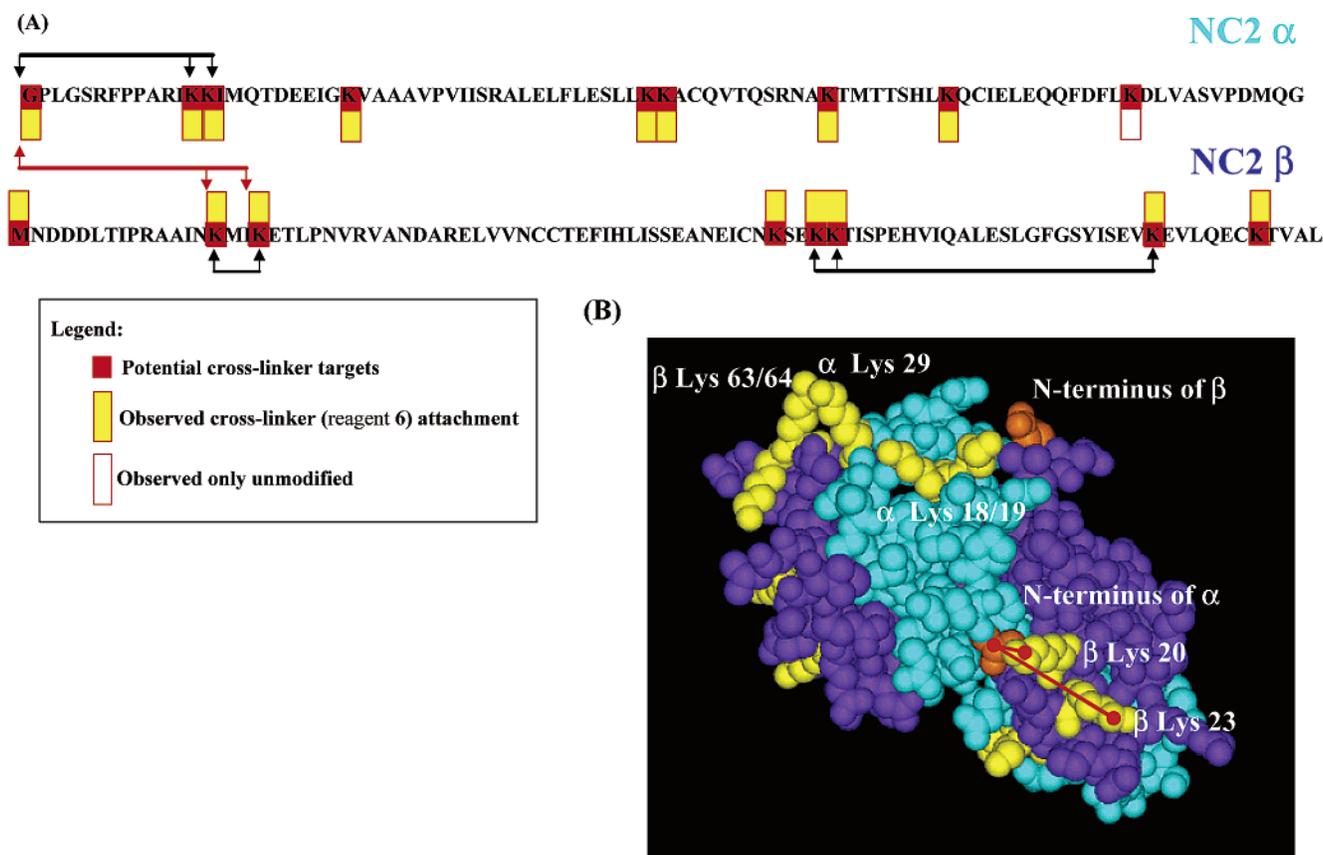


**Figure 3.** Procedure and analyses of cross-linking the NC2 complex. (A) Cross-linking procedure: (i) cross-link the protein complex; (ii) add the protein complex to the microconcentrator; (iii) wash away excess cross-linking reagent; (iv) reduce, alkylate, and then apply the endoproteinase(s); (v) add neutravidin resin (grey) and transfer to the spin column; (vi) wash away un-cross-linked digested protein fragments; (vii) cleave with  $\text{NH}_2\text{OH}$  and elute cross-linked protein fragments. (B) SDS-PAGE (Tris-glycine, 4–20%) of the cross-linking time course with **4** and the NC2 heterodimer. (C) MALDI-TOF spectra of the cross-linking time course with reagent **4** and the NC2 heterodimer ( $\alpha\beta$ ). Note: x, cross-linking reagent that attaches at both ends; y, cross-linking reagent at only one end. (D) MALDI-QqTOF of the Asp-N and trypsin digest of the NC2 heterodimer with reagent **6**. (E) MALDI-QqTOF of the Asp-N and trypsin digest post affinity purification with neutravidin beads and base cleavage of reagent **6**.

a biotin-avidin affinity purification strategy in combination with high-accuracy MS (MALDI-QqTOF) and a program (PeptideMap) for exhaustive identification of cross-linked peptides. Biotin affinity handles incorporated into our cross-linking reagents enabled us to selectively extract cross-linked proteolytic fragments and thereby reduce the complexity of the mass spectra. To bring the proteolytic fragments into the mass range of the MALDI-QqTOF ( $m/z$  4000), sequential enzymatic digestions were used. MALDI was selected for the ease of sample preparation, while the QqTOF mass analyzer was selected for its high mass accuracy (5–50 ppm). High mass accuracy is critical in reducing the number of cross-link assignments at a given nominal  $m/z$  value.

Three types of cross-links were identified: intermolecular, intramolecular, and dead-end. Intermolecular cross-links provide

information about interactions between different polypeptide chains. Intramolecular cross-links refer to connections made within a single polypeptide chain. Dead-end cross-links do not provide connectivity information but do provide information about the solvent accessibility of specific amino acid residues. Figure 4 provides a summary of these three cross-link types observed after enzymatic treatment (Glu-C, Asp-N, Asp-N/trypsin) of the NC2 complex with reagent **6**. Confident assignments of the cross-linked peptides were made when the measured mass matched a search hit unambiguously. For the most abundantly cross-linked proteolysis fragments a correlative mass shift (indicated with the arrows connecting Figure 3D to Figure 3E) was used as an additional constraint in identifying cross-linked protein fragments. When reagent **6** is used, this correlative mass shift arises from the loss of the cleaved biotin



**Figure 4.** Summary of cross-links: (A) data points identified in the cross-linking experiment of the NC2 heterodimer with reagent 6; (B) intermolecular cross-link transposed upon the crystal structure. In the space-filling model, NC2  $\alpha$  is shown in blue, NC2  $\beta$  is shown in purple, Lys residues are shown in yellow, N-termini are shown in orange, and the observed cross-link of NC2  $\alpha$  with Lys20/23 of NC2  $\beta$  is indicated in red.

arm (i.e., 456 Da). Additional information was obtained by comparing the appearance, reduction, or disappearance of mass signals prior to and after cross-linking.

To maximize MS coverage of NC2 cross-linked with the amine-reactive reagent 6, several endoproteases were applied (i.e., Asp-N, Glu-C, and the combination of Asp-N followed by trypsin). A cross-link was identified between N-terminal residues,  $\alpha$  (6–16) and  $\beta$  (6–24), using Glu-C, each containing three cross-linker reactive sites (i.e., N $^{\alpha}$ , Lys18, and Lys19 or N $^{\beta}$ , Lys20, and Lys 23). Upon digesting cross-linked NC2 with both Asp-N and trypsin, we were able to reduce the site of the cross-link to the N-terminus of  $\alpha$  (fragment residues 6–16) with either Lys20 or Lys23 of  $\beta$  (fragment residues 16–24). The observed mass of this product was 2324.15 Da (calcd 2324.24 Da). Additional information observed before and after cross-linking included the reduction and disappearance of signals, respectively, of the unmodified  $\alpha$  (6–16) and  $\beta$  (16–24) components. Formation of dead-end cross-links within these individual  $\alpha$  and  $\beta$  components was also observed.

After affinity purification, the mass spectra included large signals identified as proteolytic peptides that contained dead-end cross-linker modifications, and small signals for the intra- and intermolecular cross-links (Figure 3E). All proteolytic fragments within NC2 that contained modifiable lysine residues were observed by MS. However, signal intensities varied strongly from one peptide fragment to the next. It is noteworthy that Lys 84 of NC2  $\alpha$  was not observed with a dead-end cross-linker modification (Figure 4).

## Discussion

We have demonstrated a facile, modular strategy for synthesizing cross-linking reagents on the solid phase. With the various building blocks in hand (1–3, 7), it proved feasible to synthesize several cross-linking reagents in parallel in a single day. Modularity was demonstrated through the synthesis of five cross-linking reagents that included two or more of the following groups: an amine-reactive sulfo-NHS ester, a sulfhydryl-reactive maleimide, a photochemically reactive benzophenone, an isotope tag, a biotin affinity handle, and/or a base-labile ester cleavage site. All of the aforementioned reagents are water soluble and have attributes that improve the detectability of cross-linker containing peptides by MS. The versatility of the present synthetic procedure allows ready incorporation of other reactive groups, cleavable sites, and/or affinity handles. Thus, the researcher can optimize cross-linking reagents to fit the protein complex of interest, the question(s) of interest, and the accompanying analytical techniques.

The procedure was applied to a study of the transcription repressor NC2. The NC2  $\alpha$  and  $\beta$  components rapidly cross-linked, as predicted from the related crystal structure.<sup>16</sup> This result was confirmed both qualitatively and quantitatively, by mass spectrometry and SDS–PAGE. After enzymatic digestion, MS analysis revealed an intermolecular interaction between the N-terminal regions of NC2  $\alpha$  and  $\beta$ , as well as three intramolecular and several dead-end cross-links. The present results using the truncated NC2 construct are consistent with the X-ray

crystallographic findings of the ternary NC2–TBP–DNA complex (Figure 4).

Other interactions that might be predicted from the crystal structure but were not observed include Lys29 of  $\alpha$  with Lys63/64 of  $\beta$  and Lys18/19 of  $\alpha$  with the N-terminus of  $\beta$ . Possible reasons why these interactions were not observed include local environmental factors that may reduce reaction rates, suboptimal reaction conditions, or low MS response to certain cross-linked peptides. Failure to observe a cross-link should thus be interpreted with caution. In such cases, it may prove worthwhile to consider the application of different cross-linking reagents and/or conditions. However, in this study two different reagents were applied, giving the same result.

Advantages of the present procedure include the control afforded by the facile, modular synthetic strategy, rapid feedback using MS data at every stage, and straightforward interpretation of the resulting MS data using PeptideMap. The utilization of microconcentrators to digest cross-linked protein complexes circumvents the use of SDS–PAGE, which can exhibit smearing of cross-linked products over extended areas of the gel. Application of this integrated chemical cross-linking procedure has the potential to reveal direct protein interactions within large protein complexes and to differentiate between their conformationally distinct functional states. Finally, the use of solid-phase synthesis procedures will allow for the preparation of focused cross-linking reagent libraries. In principle, this should provide a means to optimize cross-linking studies for any given system of interest.

## Experimental Section<sup>28</sup>

**Materials.** Boc-amino acids, PyBOP, and HBTU were obtained from Novabiochem (San Diego, CA). PEGA resin was purchased from Polymer Laboratories Inc. (Amherst, MA). DMF (synthesis grade) and acetonitrile (HPLC grade) were purchased from Fisher (Pittsburgh, PA). Trifluoroacetic acid was purchased from Halocarbon (River Edge, NJ). All other reagents were obtained from Aldrich Chemical Co. (Milwaukee, WI). Sequencing grade endoproteases Glu-C (*Staphylococcus aureus* V8), Asp-N (*Pseudomonas fragi*), and modified bovine trypsin were purchased from Roche (Basel, Switzerland). Flash chromatography was performed on silica gel, 200–400 mesh (Merck).

**Reversed-Phase HPLC.** Analytical gradient HPLC was performed on a Hewlett-Packard 1100 series instrument with UV detection using a Vydac C18 column (5  $\mu$ m, 4.6  $\times$  150 mm) at a flow rate of 1 mL/min. Semipreparative and preparative gradient HPLC were performed on a Waters DeltaPrep 4000 system fitted with a Waters 486 tunable absorbance detector. Semipreparative HPLC was run on a Vydac C18 column (10  $\mu$ m, 10  $\times$  250 mm) at a flow rate of 5 mL/min. Preparative HPLC was performed on a Vydac C18 column (15–20  $\mu$ m, 50  $\times$  250 mm) at a flow rate of 30 mL/min. All runs used linear gradients of 0.1% aqueous TFA (solvent A) vs 90% acetonitrile, 10% water plus 0.1% TFA (solvent B). Low-pressure C18 columns (Michrom BioResources, Inc., Auburn, CA) were utilized for manual cleanup of the cross-linked samples.

**Mass Spectrometry.** MALDI-MS and tandem MS analyses were applied routinely to characterize all synthetic cross-linking reagents.

MALDI-MS was performed on either a MALDI time-of-flight mass spectrometer Voyager-DE STR (PE Biosystem, Foster City, CA) or a modified Sciex prototype QqTOF (Centaur).<sup>29</sup> These instruments are equipped with nitrogen lasers delivering pulses of ultraviolet light (wavelength 337 nm) at, respectively, 3 and 20 Hz to the matrix spot. Acquisitions by the MALDI time-of-flight mass spectrometer Voyager-DE STR were made in both linear and reflectron, delayed extraction modes. MALDI-TOF data were smoothed and calibrated using Data Explorer (PE Biosystem). All other spectra were smoothed, calibrated, and analyzed using the program M-over-Z (Genomic Solutions, Ann Arbor, MI). The MALDI matrixes used were 4-hydroxy- $\alpha$ -cyanocinnamic acid (4HCCA) and 2,5-dihydroxybenzoic acid (DHB). ESI-MS was performed on a commercial ion-trap mass spectrometer (Thermo Finnigan LCQ-DECA, San Jose, CA).

**Synthesis. (a) *N*-Boc-3,6-dioxaoctane-1,8-diamine.** A solution of di-*tert*-butyl dicarbonate (di-Boc) (12 g, 55 mmol, 0.5 equiv) in DCM (274 mL) was added dropwise to a mixture of tris(ethylene glycol)-1,8-diamine (15.2 g, 103 mmol, 1.0 equiv) and DIEA (10 mL, 57 mmol, 1.0 equiv) at room temperature over  $\sim$ 45 min. The reaction was stirred for 2 h more and concentrated in vacuo. Purification by flash chromatography using 0.8:0.5:8.7 methanol/propylamine/DCM as the eluent gave 6.9 g (50%) of product as an oil: <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  3.6 (s, 4H), 3.54 (t, 2H), 3.53 (t, 2H), 3.24 (t, 2H), 2.8 (t, 2H), 1.4 (s, 9H); ESI-MS (MH<sup>+</sup>) *m/z* calcd for C<sub>11</sub>H<sub>25</sub>N<sub>2</sub>O<sub>4</sub> 249.3, found 249.0

**(b) *N*-Boc-*N'*-biotinyl-3,6-dioxaoctane-1,8-diamine.** A solution of biotin (4.4 g, 18 mmol, 1.5 equiv), HBTU (6 g, 16 mmol, 1.3 equiv), and DIEA (3.5 mL, 20 mmol, 1.6 equiv) in DMF (332 mL) was stirred for 10 min at room temperature before being added dropwise to a solution of *N*-Boc-3,6-dioxaoctane-1,8-diamine (3 g, 12 mmol, 1.0 equiv). The reaction was stirred for 1 h at room temperature, after which the DMF was removed in vacuo to give an oil. The crude product was purified by flash chromatography using 0.8:0.5:8.7 methanol/propylamine/DCM as the eluent. The reaction was quantitative with a yield of 5.8 g: <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  4.5 (m, 1H), 4.3 (m, 1H), 3.6 (s, 4H), 3.54 (tt, 4H), 3.39 (t, 2H), 3.26 (t, 2H), 2.9 (dd, 1H), 2.7 (d, 1H), 2.2 (t, 2H), 1.7–1.5 (m, 8H), 1.4 (s, 9H); MALDI-QqTOF (M + Na<sup>+</sup>) *m/z* calcd for C<sub>21</sub>H<sub>38</sub>N<sub>4</sub>O<sub>6</sub>SNa 497.241, found 497.251.

**(c) *N*-Biotinyl-3,6-dioxaoctane-1,8-diamine (7).** *N*-Boc-*N'*-biotinyl-3,6-dioxaoctane-1,8-diamine (5.8 g, 12 mmol) was dissolved in 50 mL of neat TFA containing 0.5 mL of anisole and stirred for 1 h at room temperature. The TFA was evaporated under reduced pressure to give an oil which was then mixed with toluene and reevaporated to remove any residual TFA. Analytical HPLC and ESI-MS indicated the deprotection reaction was quantitative, so the crude material was used without further purification: ESI-MS (MH<sup>+</sup>) *m/z* calcd for C<sub>16</sub>H<sub>31</sub>N<sub>4</sub>O<sub>4</sub>S 375.5, found 375.0.

**(d) *N*-Succinimidyl-*N'*-biotinyl-3,6-dioxaoctane-1,8-diamine (1).** A solution containing succinic anhydride (2 g, 20 mmol, 1.7 equiv) and DIEA (10 mL, 57 mmol, 4.7 equiv) in MeCN (10 mL) was added to amine **7** (5.8 g, 12 mmol, 1.0 equiv). The reaction was allowed to proceed for around 1 h, at which point the white precipitate was dissolved and acidified with 0.1% aqueous TFA. Purification by preparative HPLC followed by lyophilization gave the desired product **1** (2.7 g, yield 48%): <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  4.5 (m, 1H), 4.3 (m, 1H), 3.6 (s, 4H), 3.55 (tt, 4H), 3.37 (tt, 4H), 3.21 (m, 1H), 2.93 (dd, 1H), 2.7 (d, 1H), 2.53 (m, 4H), 2.2 (t, 2H), 1.66 (m, 6H), 1.5 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD)  $\delta$  177.3, 176.8, 175.8, 175.6, 71.4, 70.7, 63.4, 61.7, 41.1, 41.0, 40.4, 40.3, 36.8, 31.0, 29.0, 27.0, 26.7; ESI-MS (MH<sup>+</sup>) *m/z* calcd for C<sub>20</sub>H<sub>35</sub>N<sub>4</sub>O<sub>7</sub>S 475.2, found 475.3.

**(e) *N'*-Boc-*O*-(*N*-succinimidyl-*N'*-biotinyl-3,6-dioxaoctane-1,8-diamine)serine (2).** A mixture containing **1** (238 mg, 0.50 mmol, 1.0 equiv), HATU (171 mg, 0.4 mmol, 0.9 equiv), and DIEA (0.1 mL, 0.8

(28) Abbreviations: Boc, *tert*-butyloxycarbonyl; di-Boc, di-*tert*-butyl dicarbonate; DIEA, *N,N*-diisopropylethylamine; EDC, *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide; HBTU, 2-[1*H*-benzotriazolyl]-1,1,3,3-tetramethyluronium hexafluorophosphate; HPLC, high-performance liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; NHS, *N*-hydroxysuccinimide; PEGA, poly(ethylene glycol)-poly(*N,N*-dimethylacrylamide); PyBOP, (benzotriazol-1-yl)oxytripyrrolidinophosphonium hexafluorophosphate; SPS, solid-phase synthesis; sulfo-NHS, sulfo-*N*-hydroxysuccinimide; TDA, 4,7,10-trioxatridecane-11,13-diamine. Standard IUPAC single and triple letter codes for amino acids are used throughout.

(29) Krutchinsky, A. N.; Zhang, W.; Chait, B. T. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 493–504.

mmol, 1.5 equiv) in DMF (23.4 mL) was stirred for 10 min at room temperature before being added to a solution of Boc-Ser-OH (1 g, 5 mmol, 10.0 equiv) in DMF (4.2 mL) containing a catalytic amount of DMAP. The esterification reaction was allowed to proceed at 4 °C for 5 h, at which point the DMF was removed in vacuo. The resulting oil was dissolved in 1:9 acetonitrile/water containing 0.1% TFA and purified by preparative scale HPLC to give **2** as a white solid (114.9 mg, yield 40%): <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 4.5 (m, 1H), 4.4 (s, 2H), 4.3 (m, 1H), 3.6 (s, 4H), 3.57 (t, 4H), 3.37 (t, 4H), 3.21 (m, 1H), 2.97 (dd, 1H), 2.7 (d, 1H), 2.6 (m, 4H), 2.2 (t, 2H), 1.66 (m, 6H), 1.45 (m, 2H), 1.45(s, 9H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 175.7, 175.1, 173.4, 173.2, 79.9, 71.1, 70.4, 63.4, 61.6, 54.1, 40.7, 40.3, 40.2, 36.6, 31.2, 30.5, 28.7, 28.6, 26.4; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>28</sub>H<sub>48</sub>N<sub>5</sub>O<sub>11</sub>S 662.307, found 662.301.

**(f) N-Boc-N'-succinimidyl-4,7,10-trioxatridecane-1,13-diamine (3).** Di-Boc (1.75 g, 8 mmol, 0.5 equiv) in DCM (30 mL) was added dropwise to a solution of 4,7,10-trioxatridecane-1,13-diamine (3.52 g, 16 mmol, 1.0 equiv) and triethylamine (1.2 mL, 8.8 mmol, 0.6 equiv) in DCM (40 mL). The reaction was stirred for 4 h at room temperature and then concentrated in vacuo. Purification by flash chromatography using 0.5:1:8.5 propylamine/methanol/DCM as the eluent gave 1.8 g (75%) of the product *N*-Boc-4,7,10-trioxatridecane-1,13-diamine (Boc-TDA): <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 3.62 (m, 4H), 3.59 (m, 4H), 3.5 (m, 4H), 3.1 (t, 2H), 2.7 (t, 2H), 1.8–1.7 (m, 4H), 1.5 (s, 9H).

Boc-TDA (1.1 g, 3.6 mmol, 1.0 equiv) was combined with succinic anhydride (428 mg, 4.3 mmol, 1.2 equiv) in MeCN (4 mL), pyridine (4 mL), and DCM (4 mL). The reaction was stirred for 3.5 h at room temperature before being concentrated in vacuo. Purification by flash chromatography using 0.01:1:9 acetic acid/methanol/DCM as the eluent afforded 1.3 g (91%) of the desired product. Note, trace amounts of AcOH were removed by repeated evaporation from cyclohexane (six times) followed by lyophilization: <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 3.65 (m, 4H), 3.60 (m, 4H), 3.54 (tt, 4H), 3.28 (t, 2H), 3.14 (t, 2H), 2.59 (t, 2H), 2.48 (t, 2H), 1.8–1.7 (m, 4H), 1.45 (s, 9H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 175.3, 174.2, 173.9, 158.1, 70.5, 70.4, 68.9, 37.6, 37.0, 30.7, 29.8, 29.4, 27.9; MALDI-QqTOF (M + Na<sup>+</sup>) *m/z* calcd for C<sub>19</sub>H<sub>36</sub>N<sub>2</sub>O<sub>8</sub>Na 443.237, found 443.236.

**Solid-Phase Synthesis.** All cross-linking reagents were synthesized manually according to the in situ neutralization/HBTU activation protocol for Boc solid-phase peptide synthesis.<sup>22</sup> Coupling yields were monitored by the quantitative ninhydrin determination of residual free amine.<sup>30</sup> All compounds were prepared on a Boc-Gly-[COS]-PEGA support.<sup>20</sup> The β-amino group of Dapa was orthogonally protected with the Fmoc group, which was selectively removed by treatment with 1% DBU in DMF for 1 min. Following chain assembly, cross-linking reagents were cleaved from the support by base hydrolysis (1 M NaOH, 5–20 min, room temperature), acid hydrolysis (30 mM Hg(OAc)<sub>2</sub> in 3:7 MeCN/water containing 5% acetic acid, 18 h, room temperature), or aminolysis (60 mM **7**, 195 mM thiophenol, 90 mM DIEA in DMF, overnight, room temperature). In each case, dried resin was treated with the cleavage cocktail, and the beads were filtered and washed with 0.1% aqueous TFA (buffer A), or 1:1 MeCN/water and 0.1% TFA. All compounds were purified by preparative HPLC.

**Data for compound 4:** <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O) δ 4.57 (m, 1H), 4.5 (m, 1H), 4.35 (dd, 1H), 3.9 (s, 2H), 3.63 (s, 4H), 3.6 (tt, 4H), 3.49 (dd, 2H), 3.35 (m, 1H), 3.35 (tt, 4H), 2.9 (dd, 1H), 2.7 (d, 1H), 2.62 (tt, 2H), 2.6 (tt, 2H), 2.5 (s, 4H), 2.2 (t, 2H), 1.59 (m, 2H), 1.55 (m, 2H), 1.35 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 177.6, 177.2, 175.6, 175.1, 174.8, 172.1, 70.0, 69.8, 62.7, 60.9, 55.8, 53.8, 42.6, 40.3, 39.3, 39.0, 36.1, 31.6, 30.0, 28.6, 28.3, 25.5; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>29</sub>H<sub>48</sub>N<sub>7</sub>O<sub>12</sub>S 718.308, found 718.307.

**Data for compound 5:** <sup>1</sup>H NMR (400 MHz, D<sub>2</sub>O) δ 4.57 (m, 1H), 4.51 (m, 1H), 4.35 (dd, 1H), 3.81 (s, 4H), 3.65 (s, 8H), 3.63 (s, 4H),

3.6, (tt, 4H), 3.55 (t, 4H), 3.49 (dd, 2H) 3.35 (tt, 4H), 3.35 (m, 1H), 3.2 (tt, 4H), 2.9 (dd, 1H), 2.7 (d, 1H), 2.59 (t, 2H), 2.55 (t, 2H), 2.45 (s, 4H), 2.2 (t, 2H), 1.75 (tt, 4H), 1.66 (m, 6H), 1.59 (m, 2H), 1.55 (m, 2H), 1.33 (m, 2H); MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>43</sub>H<sub>74</sub>N<sub>9</sub>O<sub>17</sub>S 1020.493, found 1020.493.

**Data for compound 6:** <sup>1</sup>H NMR (400 MHz, CD<sub>3</sub>OD) δ 4.75 (t, 1H), 4.5 (dd, 1H), 4.4 (dd, 1H), 4.37 (d, 1H), 4.32 (m, 1H), 3.9 (s, 4H), 3.63 (s, 4H), 3.55 (t, 4H), 3.37 (t, 4H), 3.21 (m, 1H), 2.93 (dd, 1H), 2.71 (d, 1H), 2.64 (dd, 4H), 2.54 (m, 4H), 2.2 (t, 2H), 1.66 (m, 6H), 1.45 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 175.3, 175.1, 175.0, 173, 171.8, 171.6, 171.2, 170.8, 73.0, 71.4, 70.4, 69.7, 64.6, 64.4, 63.1, 62.2, 61.7, 60.7, 56.9, 56.4, 56.2, 55.8, 53.6, 43.8, 41.8, 41.0, 40.9, 40.3, 39.5, 36.7, 35.8, 31.2, 30.5, 26.7, 29.7, 29.8, 29.5, 28.9, 28.6, 25.8; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>31</sub>H<sub>50</sub>N<sub>7</sub>O<sub>14</sub>S 777.322, found 777.322.

**Data for compound 8a:** <sup>1</sup>H NMR (400 MHz, DMSO) δ 8.6 (t, 1H), 8.3 (t, 1H), 8.2 (d, 1H), 7.98 (d, 2H), 7.81 (d, 2H), 7.7 (m, 3H), 7.58 (d, 2H), 6.4 (m, 2H), 4.45 (m, 1H), 4.29 (t, 1H), 4.12 (m, 1H), 3.7 (s, 2H), 3.6 (dd, 2H), 3.5 (s, 4H), 3.38 (tt, 4H), 3.18 (tt, 4H), 3.07 (m, 1H), 2.8 (dd, 1H), 2.79 (d, 1H), 2.6 (d, 1H), 2.4 (m, 4H), 2.1 (t, 2H), 1.5 (m, 4H), 1.3 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 196.3, 172.9, 170.1, 169.7, 166.9, 139.9, 138.4, 137.4, 133.8, 132.5, 129.3, 129.2, 128.3, 127.2, 70.4, 60.9, 59.1, 55.2, 42.1, 40.8, 39.5, 38.3, 34.9, 29.3, 27.9, 25.9; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>39</sub>H<sub>52</sub>N<sub>7</sub>O<sub>11</sub>S 826.345, found 826.345.

**Data for compound 8b:** <sup>1</sup>H NMR (400 MHz, DMSO) δ 8.6 (t, 1H), 8.3 (t, 1H), 8.2 (d, 1H), 7.98 (d, 2H), 7.81 (d, 2H), 7.7 (m, 3H), 7.58 (d, 2H), 6.4 (m, 2H), 4.45 (m, 1H), 4.29 (t, 1H), 4.12 (m, 1H), 3.7 (s, 2H), 3.6 (dd, 2H), 3.5 (s, 4H), 3.38 (tt, 4H), 3.18 (tt, 4H), 3.07 (m, 1H), 2.8 (dd, 1H), 2.6 (d, 1H), 2.1 (t, 2H), 1.5 (m, 4H), 1.3 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 196.3, 172.9, 170.1, 169.7, 166.9, 139.9, 138.4, 137.4, 133.8, 132.5, 129.3, 129.2, 128.3, 127.2, 69.0, 60.9, 59.1, 55.2, 42.1, 40.8, 39.5, 38.3, 34.9, 27.9, 25.0; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>39</sub>H<sub>48</sub>D<sub>4</sub>N<sub>7</sub>O<sub>11</sub>S 830.370, found 830.371.

**Data for compound 9:** <sup>1</sup>H NMR (400 MHz, DMSO) δ 8.20 (m, 1H), 8.1 (m, 1H), 7.9 (m, 1H), 7.8 (m, 1H), 7.6 (s, 1H), 7.0 (s, 2H), 6.4 (s, 1H), 6.3 (s, 1H), 4.3 (m, 2H), 4.13 (m, 1H), 3.62 (dd, 1H), 3.6 (m, 2H), 3.49 (s, 4H), 3.39 (tt, 4H), 3.19 (tt, 4H), 3.09 (m, 1H), 2.79 (d, 1H), 2.87 (dd, 1H), 2.5 (m, 6H), 2.07 (t, 2H), 1.57 (m, 2H), 1.5 (m, 2H), 1.28 (m, 2H); <sup>13</sup>C NMR (400 MHz, CD<sub>3</sub>OD) δ 175.1, 173.4, 173, 171.4, 171.2, 170.8, 170.6, 134, 69.2, 68.9, 60.9, 59.0, 55.4, 52.3, 41.0, 40.9, 40.6, 38.3, 34.9, 33.7, 32.9, 31.9, 28.0, 25.0; MALDI-QqTOF (MH<sup>+</sup>) *m/z* calcd for C<sub>32</sub>H<sub>49</sub>N<sub>8</sub>O<sub>12</sub>S 769.319, found 769.307.

**Preparation of the NC2 Heterodimer.** A truncated form of the human NC2 heterodimeric complex was obtained by coexpression of the α (residues 11–95)<sup>31</sup> and β (residues 7–99)<sup>32</sup> subunits in *Escherichia coli* (BL21[DE3]) using the expression vectors pGEX-6p-1 (Pharmacia) and pET29a (Novagen), respectively. Following cell lysis, the complex was purified from the soluble fraction by glutathione–Sephacrose chromatography. After cleavage of the GST fusion using PreScission Protease (Pharmacia), the complex was further purified by SP-Sephacrose ion exchange chromatography essentially as described.<sup>16</sup> The measured molecular weights of NC2 α, 9954 (calcd 9956.7), and NC2 β, 10460 (calcd 10462.0), confirmed that the complex was neither truncated nor posttranslationally modified during expression or purification. Gel filtration results using a Sephadex 75 column (Pharmacia) confirmed that the two subunits formed the expected heterodimer in solution (data not shown). The protein complex was stored at –65 °C as a 3 mg/mL stock solution in 10% glycerol, 200 mM NaCl, 1 mM DTT, 20 mM HEPES, pH 7, and 0.1 mM PMSF.

**Chemical Cross-Linking. (a) Activation of Cross-Linking Reagents.** Tributylammonium or cesium salts of the cross-linking reagents

(30) Sarin, V. K.; Kent, S. B.; Tam, J. P.; Merrifield, R. B. *Anal. Biochem.* **1981**, *117*, 147–157.

(31) The truncated NC2 α protein contains residues 11–95 of the native protein in addition to an N-terminal leader sequence (GPLGS).

(32) The truncated NC2 β protein contains residues 7–99 of the native protein in addition to a Met at its N-terminus.

were prepared using a cation exchange step (CM-50, Sigma). In a typical activation procedure, the corresponding cross-linker salt (25–50  $\mu\text{g}$ , 1 equiv) was dissolved in DMF (10  $\mu\text{L}$ ) along with sulfo-NHS (3 equiv), DIEA (2 equiv), and PyBOP (2 equiv). The activation was allowed to proceed on ice for 10 min, at which point the solvent was removed with a stream of argon gas using hexane as the azeotrope.

**(b) Cross-Linking Proteins with Activated Reagents.** Buffered protein (49  $\mu\text{M}$  NC2  $\alpha$ ,  $\beta$ ) was immediately added to dried activated cross-linking reagent (60–300 equiv, 4  $^{\circ}\text{C}$ ). At various time points, aliquots of the cross-linking reaction were taken for MALDI-TOF (50–100 ng of NC2  $\alpha\beta$ ), SDS-PAGE (1–2  $\mu\text{g}$  of NC2  $\alpha\beta$ ), and enzymatic digest analyses (1–5  $\mu\text{g}$  of NC2  $\alpha\beta$ ) analyses. For the photoactivatable reagent **8a**, light exposure time points were taken using a hand-held 254 nm light source. Quenching of the sulfo-NHS cross-linking reaction was accomplished with 2-mercaptoethanol followed by lowering of the pH.

**(c) Digesting and Extracting Cross-Linked Protein Complex.** To isolate cross-linked protein fragments suitable for MS analysis, we devised the procedure that is summarized in Figure 3A. NC2 previously cross-linked with reagent **4** or **6** was added to a microconcentrator (step ii), excess deactivated cross-linking reagent was washed through the filter (step iii), the protein complex was reduced with TCEP (1  $\mu\text{mol}$ , pH 7, 35  $^{\circ}\text{C}$ , 30 min), alkylated with iodoacetamide (1  $\mu\text{mol}$ , pH 7, room temperature, dark, 60 min), and digested with enzymes (Glu-C, Asp-N, Asp-N/trypsin) (step iv), biotinylated fragments were bound to neutravidin beads (250  $\mu\text{L}$  slurry, 4  $^{\circ}\text{C}$ , 30 min) and transferred to a spin column (Bio-Rad) (step v), and unmodified protein digest fragments were washed away with solutions of high salt and detergent (16 volumes, 400 mM NaCl, 46 mM *N*-octylglucoside, 50 mM Hepes) (steps vi and vii). Finally, the cross-linker-containing protein fragments

were released from the neutravidin support by cleavage of reagent **6** with hydroxylamine (200 mM, pH 8.5 at 37  $^{\circ}\text{C}$ ). MS provided a quick and informative tool for assessing each procedural step.

**Cross-Linking Search Program.** The PeptideMap program (Genomic Solutions) uses the following steps for an exhaustive calculation of all possible cross-linked and non-cross-linked peptide masses. The program (a) digests the protein sequences with the cleavage rules of the indicated specific enzyme(s), (b) calculates the peptide masses taking into account partial digestions, (c) adds the specified modifications, also taking into account partial modifications, and finally (d) calculates the masses of all possible cross-linked peptides, accounting for both complete and incomplete cross-links.<sup>33</sup>

**Acknowledgment.** We thank Sarah Keegan for assistance with the cross-linking program PeptideMap, Mike Goger for assistance with the NMR studies, Julio Camarero, Derek McLachlin, Martine Cadene, and Julio Cesar Padovan for useful discussions, and Andrew Krutchinsky for assistance with the QqTOF-MS analyses. The present work was supported in part by grants from the Burroughs Wellcome Fund, the National Institutes of Health (Grant RR00862 from the National Center for Research Resources and Grant R33CA89810 from the National Cancer Institute), the Irma T. Hirshl Trust, and the Rockefeller University.

JA026917A

- (33) Complete cross-links include those that occur between different proteolytic peptide fragments as well as those that occur within a given peptide. Incomplete or dead-end cross-links occur only within single peptide fragments.