Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 268 (2007) 217–226
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Determination of protein conformation by isotopically labelled cross-linking and dedicated software: Application to the chaperone, calreticulin

Tina Nielsen [a], Morten Thaysen-Andersen [a], Nanna Larsen [b], Flemming S. Jørgensen [c], Gunnar Houen [b], Peter Højrup [a,*]

[a] *Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*
[b] *Department of Autoimmunology, Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark*
[c] *Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark*

Received 20 February 2007; received in revised form 21 June 2007; accepted 21 June 2007
Available online 30 June 2007

Abstract

Chemical cross-linking in conjunction with mass spectrometry (MS) can be used for sensitive and rapid investigation of the three-dimensional structure of proteins at low resolution. However, the resulting data are very complex, and on the bioinformatic side, there still exists an urgent need for improving computer software for (semi-) automated cross-linking data analysis.

In this study, we have developed dedicated software for rapid and confident identification and validation of cross-linked species using an isotopic labelled cross-linker approach in combination with MS. Deuterated (+4 Da) and non-deuterated (+0 Da) bis(sulfosuccinimidyl)suberate, BS^3, was used as homobifunctional cross-linker to tag the cross-linked regions. Peptides generated from proteolysis were separated using high performance liquid chromatography, and peptide mass fingerprinting was obtained for the individual fractions using matrix-assisted laser-desorption ionisation time-of-flight (MALDI TOF) MS. The resulting peptide mass lists were combined and transferred to the program, ProteinXXX, which generated the theoretical mass values of all combinations of cross-linked peptides and dead-end cross-links and compared this to the obtained mass lists. In addition, screening for 4 Da-separated signals aided the identification of potential cross-linked species. Sequence information of these candidates was then obtained using MALDI TOF TOF. The cross-linked peptides could then be validated based on the match of the fragmentation pattern and the theoretical values produced by ProteinXXX. This semi-automated interpretation provided a high analysis speed of cross-linking data, with efficient and confident identification of cross-linked species. Four experiments using different conditions showed a high degree of reproducibility as only 1 and 2 cross-links out of 36 identified was not observed in two experiments.

The method was tested using human placenta calreticulin (CRT). Based on the identified cross-links, a few corrections to a model of calreticulin obtained by homology modelling using calnexin as template can be suggested. Furthermore, the cross-links show that the C-terminal of the protein continues along the core region opposite the P-domain for at least 11 residues beyond the known structure. In addition, it was observed that the conformation of CRT does not change significantly in the presence or absence of the divalent ions, Ca^{2+} and Zn^{2+}.

Keywords: Cross-link; Isotope labelling; Mass spectrometry; Three-dimensional protein structure; Calreticulin

1. Introduction

As a result of the genome sequencing efforts, a substantial number of novel proteins have been identified during the last decade. The structures of these proteins have mainly been investigated on the sequence level and the higher orders of structures often remain unsolved. The functions of proteins are closely linked to their three-dimensional (3D) structures and the presence of interaction partners. Thus, the methods for elucidation

Abbreviations: 3D, three-dimensional; BS^3, bis(sulfosuccinimidyl) suberate; CNX, calnexin; CRT, calreticulin; ER, endoplasmatic reticulum; MALDI, matrix-assisted laser desorption/ionisation; MeCN, acetonitrile; MS, mass spectrometry; MS^2, tandem mass spectrometry; NMR, Nuclear magnetic resonance; SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis; TFA, trifluoroacetic acid; TOF, time-of-flight

* Corresponding author. Tel.: +45 6550 2371; fax: +45 6550 2467.
E-mail address: php@bmb.sdu.dk (P. Højrup).

1387-3806/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2007.06.019

of spatial and topological organization of proteins have attracted much attention.

Two methods are mainly used for high resolution determination of 3D structures: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. However, both techniques are insensitive and require large amounts of pure proteins. Chemical cross-linking of proteins is a low-resolution technique that can provide information about tertiary and quaternary structure. With the advent of mass spectrometry (MS), this technique has experienced a major revival [1–3] and cross-linking approaches now feature some advantages, which, for some experiments, make it an attractive alternative. These advantages include high sensitivity, rapid analysis time, independency of protein mass (using bottom-up approach) and a possibility to analyse post-translationally modified proteins. In addition, the specificities of the cross-linking reagents to the functional groups of the amino acid residues are very broad, resulting in a wide variety of experiments that can be performed [4]. The major disadvantage of using chemical cross-linking arises from the fact that reaction mixtures tend to become highly complex, in particular when dealing with large proteins. Thus, approaches have been developed to enrich cross-linked products by affinity chromatography [5,6] or to facilitate the identification of cross-linked species using isotope [7–9] or fluorescence-labelled cross-linkers [10]. Additionally, some chemical or mass spectrometric cleavable cross-linkers have been used to ease detection of the cross-linked products [11–13].

As pointed out by Sinz in a recent review [14], one of the major limitations for employing chemical cross-linking and MS analysis for rapid protein structure characterization is the lack of computer software that efficiently can analyze the extreme complexity of reaction mixtures. Several software programs for identifying chemical cross-links are freely available on the Web (see [14] for further description of the individual programs): Automated Spectrum Assignment Program (ASAP), MS2Assign, MS2PRO [2,9,15,16] (all available at http://roswell.ca.sandia.gov/~mmyoung/) and SearchXLinks software (available at http://searchxlinks.de/cgi-bin/home.pl) [17,18]. However, all the existing programs have specific limitations, and time-consuming manual assignments and validations of cross-linked peptides are often necessary.

Calreticulin (CRT) is, together with its homologous partner, calnexin (CNX), a molecular chaperone involved in protein folding and assembly. These endoplasmic reticulum (ER)-resident lectins bind soluble and membrane-bound glycoproteins synthesized into this compartment, and it has been shown that they specifically associate with glycoproteins bearing monoglucosylated high-mannose oligosaccharides ($Glc_1Man_9Glc_1NAc_2$) [19].

At present, no overall 3D structure of CRT is available. However, based on crystallographic studies of CNX [20] and NMR studies of the P-domain of CRT [21], a 3D model was proposed [22]. The N- and P-domains of CRT are likely responsible for the chaperone function of the protein, whereas the acidic C-domain is involved in high capacity Ca^{2+} binding and likely plays an important role in calcium homeostasis [23]. The proline-rich P-domain forms an extended arm structure and associates with the thiol oxidoreductase ERp57, which promotes disulfide formation/isomerization in glycoproteins [24].

Alterations in the ER environment, such as changes in the Ca^{2+} and Zn^{2+} concentrations, have been suggested to affect the conformation of CRT and thus change its ability to assist in protein folding [25]. Together with CNX, CRT and ERp57 comprise the CRT/CNX cycle, which is believed to be responsible for the folding and quality control of newly synthesized glycoproteins. Substrates bind transiently with CNX and CRT, and may undergo repeated cycles of deglucosylation/reglucosylation before they are properly folded [26].

In this study, we developed a novel bioinformatic tool for cross-linking analysis that can utilize an isotope-labelled cross-linker, and we have used this technique to investigate the 3D structure of human placenta CRT. Additionally, we used this approach to study the influence of Ca^{2+} and Zn^{2+} on the overall conformation of CRT.

2. Experimental

2.1. Purification of calreticulin

CRT was purified from a human placenta essentially as described [27]. Human placenta was homogenized twice at 4 °C with 20 mM bis-Tris, pH 7.2, and twice with 20 mM bis-Tris, pH 7.2, 1% (v/v) Triton X-114. The supernatant and the precipitate were separated between extractions by centrifugation at 16,300 × *g* for 1 h at 4 °C. The supernatants were pooled and diluted with an equal volume of 20 mM bis-Tris, pH 7.2, 3% Triton X-114 and incubated at 37 °C overnight to separate the water and detergent phases. Ammonium sulphate (337 g/l) was added to the water phase. Following an overnight stirring at 4 °C, the precipitate was removed by centrifugation at 16,300 × *g* for 1 h at 4 °C and the supernatant was subjected to ultradiafiltration against 6 volumes of 20 mM Tris, pH 7.5, using a 10 kDa cut off filter. This fraction was then applied to a Q Seqharose ion-exchange column (2.6 cm × 33 cm) equilibrated with 20 mM Tris, pH 7.5. The protein was eluted using an gradient of increasing concentration of NaCl (0–0.45 M) in the same buffer. The CRT-containing fractions were identified by SDS-PAGE, immunoblotting and immuno assays using antisera recognizing the C-terminal of CRT. The relevant fractions were pooled, concentrated by ultradiafiltration against 20 mM Tris, pH 7.5, and subsequently purified by size-exclusion chromatography on a sephacryl S-100 HR column (2 cm × 78 cm). The eluting fraction showed a single band of approximately 60 kDa in SDS-PAGE.

2.2. Experimental setup and cross-linking

For each of the following experiments, 2 nmol purified CRT in 100 μl 20 mM Tris, pH 7.5 was desalted into 0.1 M PBS pH 7.8 using a Fast Desalting PC3.2/10 (Sephadex G-25) column (Amersham Pharmacia Biotech, Uppsala, Sweden) and dried to a final volume of 200 μl.

Three separate CRT cross-linking experiments were carried out (final concentrations are listed): (I) presence of 1 mM Ca^{2+},

(II) presence of 1 mM Zn^{2+} and (III) depletion of divalent ions (5 mM EDTA). The samples were left at 37 °C for 30 min to allow changes in conformation.

The cross-linking was carried out at 50-fold molar excess of the cross-linker by adding 50 nmol of bis(sulfosuccinimidyl)suberate (BS^3)-d_0 and BS^3-d_4 both in 1 μl DMSO (Pierce Biotechnology, Rockford, IL) to the samples, which were incubated for 90 min at *RT*. The reaction was terminated by the addition of NH_4HCO_3, pH 7.8 to a final concentration of 50 mM.

2.3. Proteolytic digestion and HPLC

The samples were enzymatically digested for 2 h at 37 °C using 1.5% (w/w) of modified trypsin (Promega, Madison, WI). Subsequently, the same amount of trypsin was added, and the samples were incubated overnight at 37 °C. To ensure optimal digestion conditions of trypsin, a final concentration of 6 mM $CaCl_2$ was added to the EDTA containing sample (III). Similarly, a control sample, containing 2 nmol of non-cross-linked CRT was digested as an HPLC reference.

Following digestion, 0.5 mM DTT was added to the mixture in order to reduce disulfide bridges. After 30 min at 37 °C the peptides were separated by reversed phase chromatography on a Jupiter C_{18} 250 mm × 2 mm, 5 μm, 300 Å column (Phenomenex, Torrance, CA). The separations were carried out on an Äkta-Basic or Ettan HPLC (Amersham Pharmacia Biotech, Uppsala, Sweden) at a flow rate of 250 μl/min. The gradient, made of B buffer (0.05% trifluoroacetic acid (TFA) + 90% acetonitrile (MeCN) in water), increased from 5 to 45% in 40 min; 45 to 60% in 5 min; 60 to 90% in 2 min. The A buffer consisted of 0.06% TFA in water. UV-absorbing fractions were collected manually, as well as fractions in between peaks, dried and redissolved in 20 μl 0.1% TFA.

From some fractions containing large (cross-linked) peptides a few microliters were redigested at 37 °C for 2–3 h using 0.05 μg endoproteinase Asp-N (Merck, Darmstadt, Germany) in 20 μl 50 mM NH_4HCO_3, pH 7.8.

2.4. Sample preparation and mass spectrometry

Five microliters of each sample (trypsin ± endoproteinase Asp-N digested) were purified using hydrophobic microcolumn material and loaded directly onto the target. GelLoader pipette tips (Eppendorf, Hamburg, Germany) were partially constricted by squeezing the end and a 5 μl slurry (Poros R2, 20 μm, Applied Biosystems, Framingham, MA) was loaded onto the tip and packed by applying air pressure with a 1 ml syringe. The resin (2–5 mm in height) was washed with 100% MeCN and equilibrated by flushing 20 μl 0.1% TFA through the column. The sample was loaded onto the column and washed in the same buffer. The sample was eluted onto the target by 0.8 μl of matrix solution containing 5 mg/ml α-cyano-4-hydroxycinnamic acid in 0.1% TFA, 70% MeCN.

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF TOF MS) was performed using a 4700 Proteomics Analyzer (Applied Biosystems, Framingham, MA). Peptide mass fingerprinting of all the samples was obtained using automatic MS acquisition and the external calibration of the instrument was made in between samples using a tryptic digest of lactoglobulin (bovine). A combined peak list was obtained using Data Explorer Ver 4.5 (Applied Biosystems, Framingham, MA) and MSfileConvert and imported into ProteinXXX (both freely available on www.gpmaw.com), where the candidate cross-linked peptides were identified. Sequence information was obtained on these potentially cross-linked species using MALDI TOF TOF MS^2 and validated by comparing the MS^2 data to the theoretical values provided by ProteinXXX.

2.5. Modelling of calreticulin

The sequence of CRT was extracted from the Swiss-Prot database (P27797). As template structure for the homology modelling, the structure of CNX was used [20]. Coordinates of CNX were taken from the Protein Data Bank (1JHN). Sequence alignment was performed with the ClustalX program [28] and subsequently manually adjusted. The 3D model of CRT was based on the sequence alignment to CNX shown in Fig. 1. For two regions on 11 and 7 residues, respectively, the structure of CNX have not be determined, and accordingly no template for the model building of CRT was available in these two regions (Fig. 1, regions A and B). The Modeller 8v1 program by Sali and Blundell [29] was used for the homology modelling of CRT from CNX. The initial homology model of CRT was extended towards the C-terminal region. The extension was originally constructed as an ideal alpha-helix, and subsequently the backbone torsional angles were manually adjusted to avoid steric clashes between the original part of the CRT structure and the C-terminal extension. Thus, the C-terminal part of the CRT model is not based on any template and accordingly it has only been included for illustrative purposes. The stereochemical quality of the model structures were checked by PROCHECK accessed via the ADIT validation server [30].

2.6. MSfileConv and ProteinXXX

MSfileConv and ProteinXXX were written in Borland Delphi 2006 and compiled to individual executable files. MSfileConv reads individual peak lists in ASCII format generated by the Data Explorer software and combines the files to a single peak list where each list is headed by a line 'TITLE = file name'. An option enables multiple occurrences of a mass value to be collapsed into a single averaged value.

ProteinXXX is a module from GPMAW (Lighthouse data, Odense, Denmark) which has been extracted as a separate program and made available as freeware.

Protein sequences can be read into ProteinXXX in various formats (GPMAW, FastA, EMBL and GenPept) and they can be pasted and edited from additional sources.

Analysis of cross-linking takes place through a user-friendly wizard. First, the user defines the protein sequences to participate in cross-linking. These have to be loaded into the program beforehand, and cross-linking can be defined between two identical or different proteins.

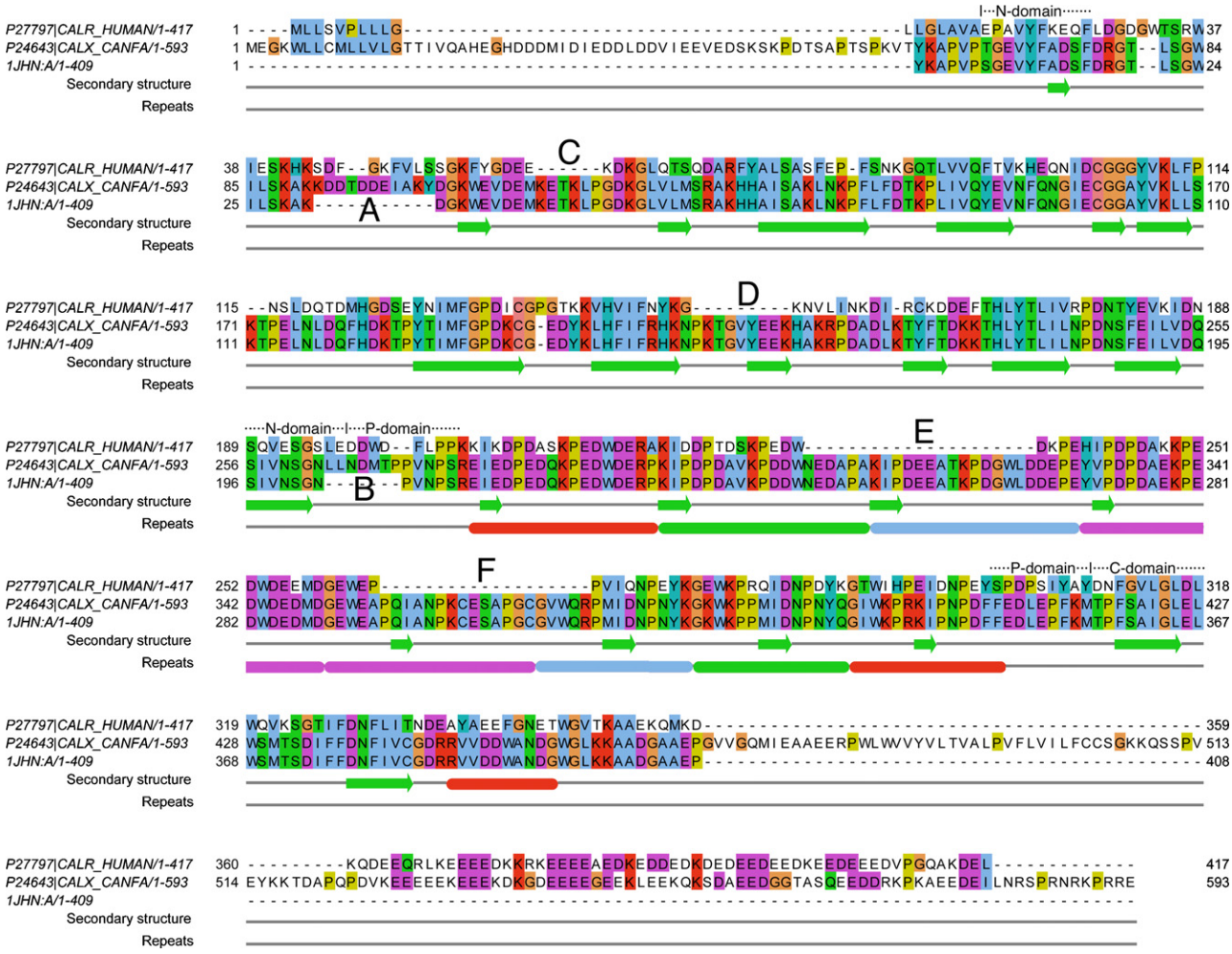

Fig. 1. Sequence alignment of the full sequences of CRT (Swiss Prot entry P27797), CNX (Swiss Prot entry P24643) and the sequence of the experimentally determined structure of CNX (PDB entry 1JHN). The residues are coloured according to the standard ClustalX colour code. Below the sequences the secondary structure of CNX has been displayed as green arrows (beta-sheets) and red bars (alpha-helices), respectively. The locations of the 4 × 2 repeats present in the CNX structure have been display by colored bars below the secondary structure assignments. The letters A–F refer to residues not observed in the CNX structure (A and B), loops shorter in CRT than in CNX (C and D), and the missing repeats in CRT (E and F). Finally, the domain borders for CRT are marked above the alignment.

Second, define the cross-linker to be used along with the enzyme(s) used for cleavage, number of missed cleavage sites, and an upper limit for mass analysis. Either zero-length, homo-bifunctional or hetero-bifunctional cross-linkers can be used. The cross-linkers are defined in a user-editable database with a tabular interface. For each cross-linker, the user has to define the composition of the linker in cross-linking mode, hydrolyzed and reduced form (if applicable). In addition, linkage type and chemical groups to link have to be specified.

Finally, a window opens with calculated mass values for peptides, two cross-linked peptides, peptides with one or more attached dead-end (hydrolyzed) linkers, cross-linked peptides with a dead-end linker and peptides with an internal cross-link. The various linkage types can be selected and deselected individually.

The MS^2 fragmentation pattern for any peptide can be displayed, either by double-clicking on the appropriate line or by selecting the MS^2 button. The linked peptides are shown graphically, and if multiple linkage options are possible (cross-links and/or hydrolyzed linkers), drop-down selection boxes enable displaying the different options.

In order to search for possible cross-links, a mass list of at most 500 mass values can be imported, either from disk or pasted into a table. Mass lists larger than 500 mass values can be read from disk in a proprietary format that can be easily adapted from other formats (see above). The mass search is performed with a user-defined precision (in parts per million) and the results are shown on a separate page. Here, the results are shown as mass values, peptides, linkage type and accuracy. Each ‘hit’ can be shown as MS^2 fragmentation in the same way as the primary list.

When performing a mass search it is possible to specify an isotope pattern difference (up to 8 Da) where the search program will extract all mass values with the given difference i.e., ±0.1 Da. The extracted list will then be searched against the cross-link list. The total result from this search (isotope pattern and cross-link hits) is shown on a separate page with options like for the normal hit.

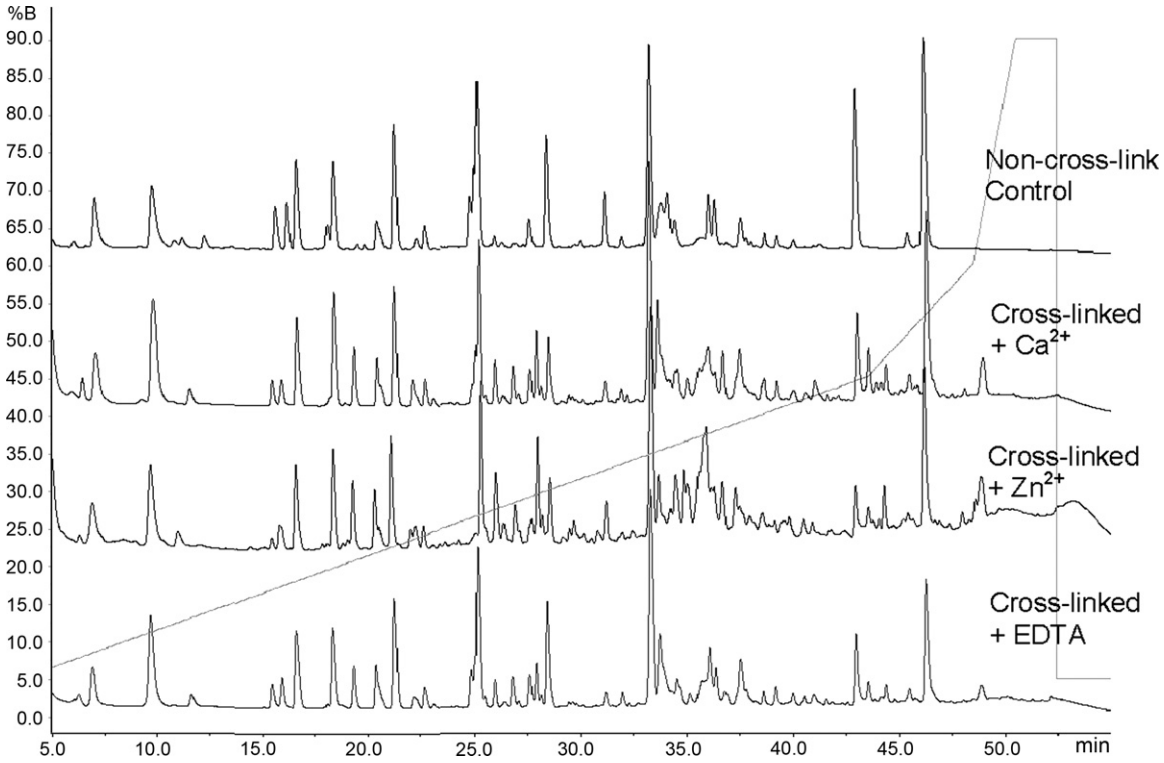

Fig. 2. Comparison of peptide UV_{214} elution profiles using C_{18} reversed-phase chromatography of the four samples investigated in this study (top to bottom): CRT without cross-linker (control), CRT with cross-linker and Ca^{2+}, CRT with cross-linker and Zn^{2+} and CRT with cross-linker and EDTA.

Individual searches can be saved and compared in a table in order to obtain and locate identical and non-identical hits in up to six experiments.

3. Results/discussion

In order to analyse the three-dimensional (3D) structure of the human chaperone calreticulin (CRT), we have developed software tools and a workflow to analyse the large amount of data generated during cross-linking experiments with mass spectrometric detection. The 3D structure of CRT is unknown, but a model has been built based on the closely homologous protein calnexin (CNX). However, the construction of the model is complicated by the fact that CNX is considerably larger and an alignment of CRT and CNX shows a number of rather large deletions in CRT [31] (Fig. 1, regions C–F). These deletions correspond to loop structures and beta strands situated at the edge of the CNX core structure, hence the overall structure of CRT is still believed to be highly similar to that of CNX. Furthermore, for two regions on 11 and 7 residues, respectively, the residues in CNX could not be determined and accordingly no template was available for these regions (Fig. 1, regions A and B). An additional goal was to map the C-terminal part of CRT, as this is not part of the published CNX structure and therefore not included in our initial homology model of CRT.

Initially, the cross-linking experiment was carried out in the presence of a small amount of Ca^{2+} (1 mM) to investigate the native 3D structure of CRT. Deuterated (+4 Da) and non-deuterated bis(sulfosuccinimidyl)suberate, BS^3, was used as a homobifunctional cross-linker to tag the cross-linked regions. Proteolytic digestion, using trypsin, generated a complex mixture of peptides, which was then separated using reversed-phase high-performance liquid chromatography (Fig. 2, trace 2). A total number of 46 fractions were collected and analysed using matrix-assisted laser desorption ionisation time-of-flight (MALDI TOF) MS in an automated manner. The resulting peak lists of the individual fractions were combined using MSfileConv (total of 1805 mass values) and imported into ProteinXXX. In order to generate a list of all possible cross-linked peptides, human Calreticulin (Swiss-Prot P27797) was loaded, and the 17 N-terminal residues were removed (corresponding to the signal peptide). Using parameters of five missed cleavages, upper mass limit of 5000 Da and the standard definition of BS^3, ProteinXXX generated 7292 possible entries of peptides, cross-linked peptides, peptides with dead-end linker(s) (one end of the linker hydrolyzed) and cross-linked peptides with one dead-end linker. Homopeptides were omitted from the search as SDS-PAGE gel analysis showed that CRT dimers were not present (data not shown). By comparing this list of theoretical masses to the observed values with a precision of 50 ppm, ProteinXXX returned a list of potential candidates (containing 1165 entries). The list of candidates was further screened for 4 Da separated signals, which reduced the list to 281 values when only counting +4 Da signals originating from the same fraction. The list could be further reduced by using the built-in graph for displaying the accuracy of the +4 Da signal, which showed a clear clustering of potential linkers (within ±0.05 Da) with outliers that in no case could be verified as actual linked peptides. As multiple candidates often matched the same specific mass value, it was essential to acquire further sequence information for con-

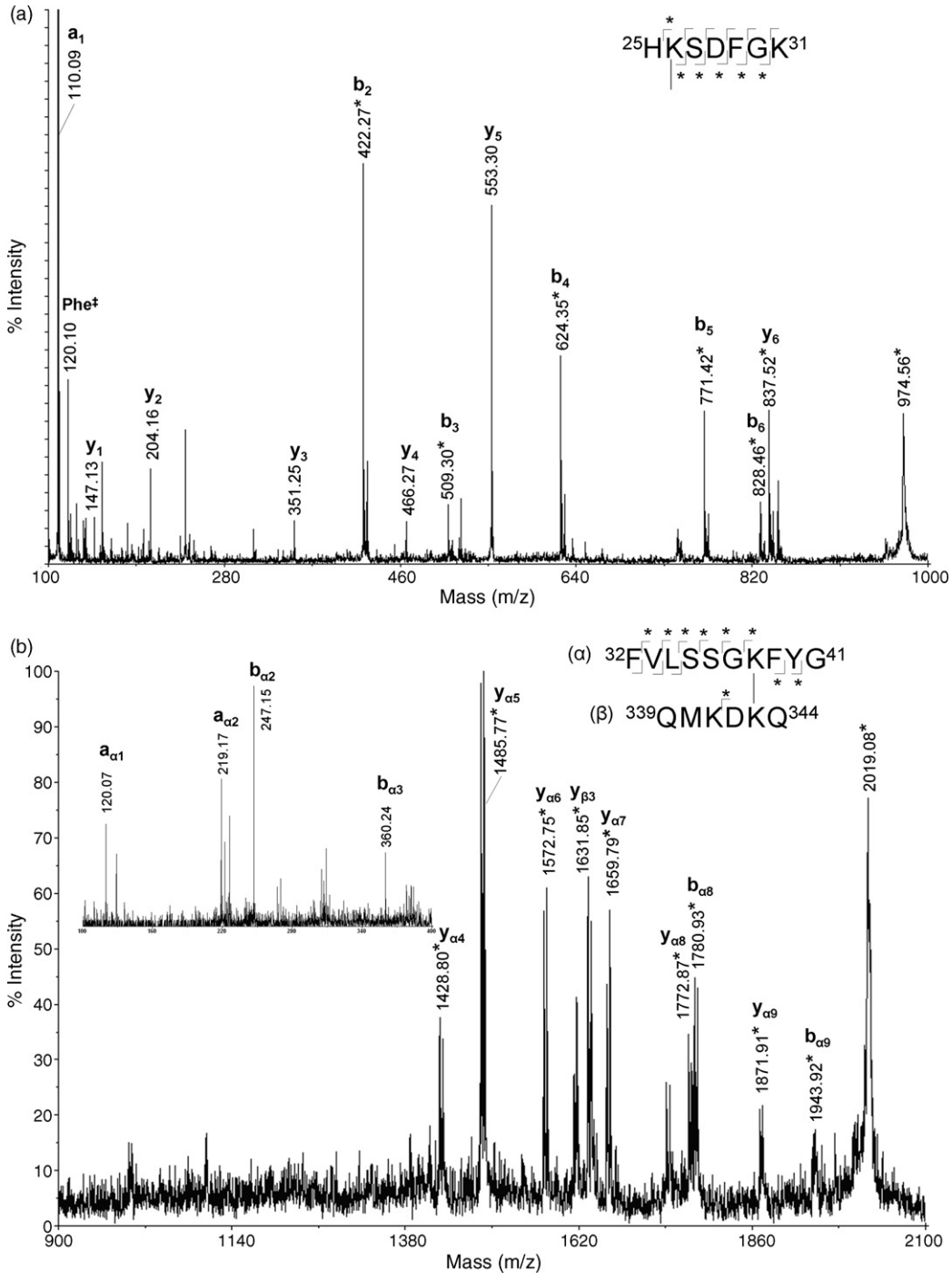

Fig. 3. MALDI TOF TOF MS^2 fragmentation of (A) a dead-end (hydrolyzed) cross-linker bound peptide and (B) an inter-peptide cross-linked product. The identified fragment ions and the validated peptides are presented (inserts). *The given fragment ion was observed with +0 Da and +4 Da. Residues marked with ‡ indicate immonium ions.

fident validation. This was obtained using MALDI TOF TOF MS^2. The MS^2 data were usually obtained using the same spot set used for the initial mass scouting. Fig. 3 presents examples of validations of a dead-end cross-linked peptide and an inter-peptide cross-linkage using MS^2. In addition, some orthogonal information was obtained for a few cross-linked species using an alternative proteolytical enzyme (e.g., endoproteinase Asp-N) in combination with MS.

Table 1 lists the observed BS^3-modified species in the Ca^{2+}-treated CRT, including dead-end linked peptides and intra-/inter-peptide cross-links. Altogether, 21 dead-end, 10 intra-peptide cross-links and 5 inter-peptide cross-links were identified. In addition, several missed cleavage variants of these peptides were found, and several lysine residues were involved in multiple types of cross-links. To illustrate the localization of the observed cross-linked species, these are mapped on the primary sequence

Table 1
Dead-end (hydrolyzed), intrapeptide and interpeptide cross-linkers identified in the present work

Cross-link type	Peptide from-to	Theoretical *m/z*	Observed *m/z*	Linked residue	Sequence
Dead-end (hydrolyzed) linker					
	1–7	1009.52	1008.55	1	**E**PAVYFK
	20–26	1083.58	1082.62	24	WIES**K**HK
	25–31	974.49	974.49	26	H**K**SDFGK
	27–38	1427.74	1427.76	31	SDFG**K**FVLSSGK
	32–47	2004.98	2005.02	38	FVLSSG**K**FYGDEEKDK
	39–47	1286.58	1286.59	45	FYGDEE**K**DK
	46–56	1374.69	1374.71	47	D**K**GLQTSQDAR
	68–81	1704.05	1703.99	70	SN**K**GQTLVVQFTVK
	126–134	1303.74	1302.77	126	**K**VHVIFNYK
	127–136	1360.76	1360.79	134	VHVIFNY**K**GK
	137–145	1240.73	1240.73	142	NVLIN**K**DIR
	146–168	2955.45	2955.77	147	C**K**DDEFTHLYTLIVRPDNTYEVK
	169–190	2675.27	2675.28	189	IDNSQVESGSLEDDWDFLPP**K**K
	190–205	2241.09	2240.08	190/192	**K**I**K**DPDASKPEDWDER
	193–205	1715.74	1715.75	198	DPDAS**K**PEDWDER
	206–231	3271.55	3271.52	207/221	A**K**IDDPTDSKPEDWD**K**PEHIPDPDAK
	256–261	928.49	928.51	259	GEW**K**PR
	318–338	2500.13	2500.21	334	DEAYAEEFGNETWGVT**K**AAEK
	342–349	1203.54	1203.56	343	D**K**QDEEQR
	350–359	1459.76	1459.78	351	L**K**EEEEDKKR
	350–359	1459.76	1458.82	358	LKEEEEDK**K**R
Intrapeptide cross-link					
	1–19	2383.12	2383.16	1–7	**E**PAVYF**K**EQFLDGDGWTSR
	127–142	2024.17	2024.15	134–136	VHVIFNY**K**G**K**NVLINK
**	146–155	1408.61	1408.64	146–147	**C**KDDEFTHLY
	169–192	2898.44	2898.42	189–190	IDNSQVESGSLEDDWDFLPP**KK**IK
	190–205	2067.00	2067.00	190–192	**K**I**K**DPDASKPEDWDER
	191–205	1938.91	1938.93	192–198	I**K**DPDAS**K**PEDWDER
	335–349	1971.94	1971.99	338–341	AAE**K**QM**K**DKDEEQR
	342–357	2186.01	2186.01	343–351	D**K**QDEEQRL**K**EEEEDK
	350–358	1285.65	1285.65	351–357	L**K**EEEED**K**K
	350–359	1441.75	1441.78	357–358	LKEEEED**KK**R
Interpeptide cross-link					
	1–19, 39–47	3512.62	3512.50	7–45	EPAVYF**K**EQFLDGDGWTSR, FYGDEE**K**DK
**	25–31, 46–56	2174.08	2174.08	26–47	H**K**SDFGK, D**K**GLQTSQDAR
	32–41, 339–344	2018.25	2018.00	38–343	FVLSSG**K**FYG, QMKD**K**Q
*	126–142, 190–192	2539.55	2539.34	134/136–190	KVHVIFNY**K**G**K**NVLINK, **K**IK
	342–349, 350–358	2332.12	2332.18	343–351	D**K**QDEEQR, L**K**EEEEDKK

The residue to which the linker is attached is marked in bold. In the case of peptide 190–205, we were unable to determine whether residue 190 or 192 was attached to the linker. All peptides were observed in all experiments except the peptides marked with *, not observed in Zn^{2+} experiment, and **, not observed in Ca^{2+} experiment. The theoretical *m/z* values include the mass of a hydrolyzed linker (138.068 Da) or cross-linker (156.079 Da), respectively.

of CRT, Fig. 4. In addition to the cross-linked peptides, most of the non-modified peptides were observed in the MS analysis, the only exception being the C-terminal peptides from residue 374.

Almost exclusively, only lysine residues and the N-terminus of CRT were observed to be modified by BS^3. However, Cys146, which is known to be surface exposed and in the reduced form in active CRT [32], is part of a cross-link. As Cys146 was not observed with a dead-end linker, this indicates that modification of Lys147 increased the local concentration of the SH-group sufficiently for a reaction to take place. No modification of the disulfide bonded Cys88–Cys120 residues was observed.

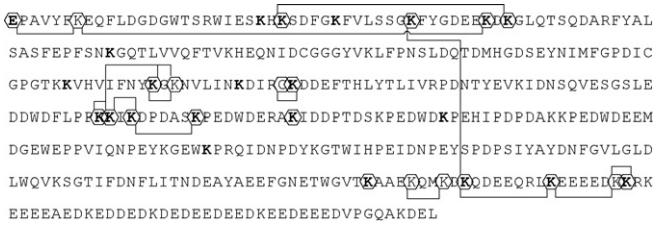

Fig. 4. Sequence of human CRT in 1-letter code, illustrating the observed linkers listed in Table 1. Dead-end linkers are shown in bold letters; intra- and interpeptide links are drawn with a line. Most residues participating in cross-linking were also found with dead-end linkers.

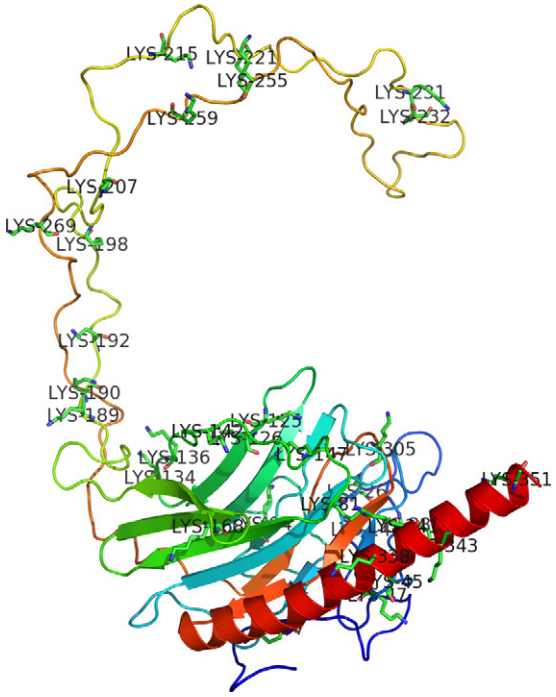

Fig. 5. Three-dimensional structure of the model of CRT based on the known structure of CNX and extended with the results obtained in this study. Only the backbone is drawn except for lysine residues, which are drawn as sticks. The figure was created in Pymol [36].

In several cases, the observed mass of peptides with a dead-end linker was 1 Da less than the theoretically calculated mass. As the peptides were all validated by MS^2 analysis, we are confident of the identification. As the cross-linking reaction was terminated by the addition of ammoniumbicarbonate the reason for the 1 Da difference is most likely that some free cross-linkers are not hydrolyzed, but react with ammonia instead, resulting in a 1 Da smaller peptide mass. In several cases, this could also be observed in the isotopic pattern of the BS^3-modified peptide, which showed a +1 peak higher than expected relative to the monoisotopic peak (data not shown), indicating that part of the linker had reacted with ammonia and part had reacted with water. This shows that the cross-linking reaction under the present reaction conditions is not terminated even after 90 min, and thus excludes the method for probing rapid changes in protein conformation.

Based on the observed cross-linked species mainly from the Ca^{2+} treated CRT (Table 1), some important information about its 3D structure could be deduced by mapping the cross-links to a 3D model, Fig. 5. This model is based on the published structure of CNX (PDB database 1JHN) and clearly divides the protein into three domains: Glu1-Pro188 makes up the N-domain (core region) in the form of a two-layer beta-sheet. Lys189-Ser287 creates an elongated arm, stabilized by short regions of beta-sheet. Ile288-Glu319 winds through the core structure as two beta strands forming an integral part of the core domain. From Ala320 there is an alpha helix positioned on the surface of the core region. From the last residue in the initial model, Val332, the remainder of the strongly acidic C-terminal region, is missing.

A minority of the amino groups in the protein (14 out of 43 lysines including the N-terminus) was not observed modified by BS^3. The reason for not observing these modifications is likely that five of them are located in the very acidic C-terminal region. Hence, these are localised in very large tryptic peptides, that are difficult to detect by MS [32] and the remainder will be present in large peptides due to missed cleavages.

When cross-linking two amino groups, BS^3 introduces a spacer arm with a length of 11.4 Å. Measuring the distances between the epsilon amino groups of lysine residues in the model leads to the following observations: in the intrapeptide link observed in peptide 1–19, the link is between the N-terminus and Lys7. In the model, this distance is 6–7 Å, but as the N-terminus is likely to be rather flexible, it can be accommodated in the model. The interpeptide linkage observed between Lys7 and Lys45 has a distance in the model of 19 Å, and the residues are situated on opposite sides of the structure. Although this distance depends heavily on the lysine side-chain torsional angles and therefore in principle could be considerable smaller, it is more likely that the first beta strand must be shifted or, more likely, the turn, which Lys45 is part of, is incorrectly modelled. The Lys45 is part of a loop in CRT being considerable shorter than the corresponding loop in the CNX template structure (Fig. 1, region A). The next cross-link between Lys134 and Lys136 is 13 Å, which also agrees with the model. The rather unusual linkage between Cys146 and Lys147 indicates that either one or both of the residues need to be twisted slightly, but this should not change the model significantly. The next linkage Lys134/Lys136–Lys190 is between the N-domain and the P-domain. Based on the mass spectra, it was not possible to determine whether linkage was to Lys134 or to Lys136, which are situated closely together in the model. However, Lys190 is located on the opposite side of the P-domain facing away from the core. This part of the P-domain thus needs to be twisted to bring Lys190 to the other side, where it can interact with Lys134 and/or Lys136. Both Lys134/Lys136 and Lys190 are located close to regions where the template for the model building is less reliable (Fig. 1, regions D and B, respectively). As the three next links all involve the same region of the P-domain (e.g., Lys189–Lys190, Lys190–Lys192 and Lys192–Lys198) and are compatible with the linker in the present model, the whole region needs to be twisted to face the core region. This could also explain why tryptic domain mapping of calreticulin [32] failed to show cleavage in the P-domain, as all the lysine residues would have been shielded against the core region.

Finally, in the C-terminal domain, a link was found between Lys38 and Lys343. This shows that at least the first part of the C-terminal domain is in close contact with the core region. The distance from the C-terminal of the model is approximately 29 Å. This is within easy reach if this part of the structure is in extended conformation, but also an alpha-helix is possible, particularly as Lys38 is part of the loop structure which also includes Lys26 and which, according to the linkage with Lys47, is likely to be shifted considerably closer to the N- and C-termini. In support of the alpha-helical theory is the finding of Lys338–Lys341 and Lys343–Lys351 linkages. With a distance of 3 and 8 residues they fit with the 3.6 residues/turn observed in the alpha helix.

Secondary structure predictions for CRT (not shown) indicated that the C-terminal alpha-helix observed in the CNX structure perhaps extended further. For illustrative purposes the homology model was manually modified to include an extended alpha-helical C-terminal. The following linkages of Lys351–Lys357 and Lys357–Lys358 are more ambiguous and may fit better with beta- or turn-like structures.

In several reports [33–35] divalent cations have been implemented in functional and structural changes in CRT. In order to investigate this influence, we treated CRT with Ca^{2+}, Zn^{2+} and EDTA for depletion of divalent cations in separate cross-linking experiments identical to the one described above. The chromatographic profiles of these samples (including a non-cross-linked control) were compared, Fig. 2. The cross-linked CRT samples shared some similarities with the non-cross-linked CRT sample. However, several additional peaks appeared in these chromatograms due to the increased complexity caused by the cross-linking. High similarity of the elution profile of the Ca^{2+} and Zn^{2+} treated sample was observed, although some minor differences appeared in the latter part of the chromatograms. Additionally, the ion depleted cross-linked CRT sample gave rise to a highly similar elution profile. MS analysis was performed on all fractions and the same set of cross-linked species as for the Ca^{2+} treated CRT was observed, showing that the experiment was reproducible although intensities for the individual products varied slightly (data not shown). The reproducibility of the method is further affirmed by the fact, that the two $Ca^{++} \pm$ EDTA and $Zn^{++} \pm$ EDTA experiments were carried out several month apart. Only the 146-155 intrapeptide and the 25-31/46-56 interpeptide were not found in the Ca^{2+} enriched sample and the 190-192/126-142 interpeptide was not found in the Zn^{2+} enriched sample. As no other linkages were observed instead, it indicates that the presence/absence of divalent ions does not affect the structure of CRT in major ways. However, as we did not observe peptides in the last part of the C-domain (i.e., after residue 360) there is still the possibility of major changes here, particularly as this part is believed to bind a large number of Ca^{2+} in the native molecule.

Since an isotopically labelled cross-linker was used, the above-mentioned experiments could alternatively be performed in a quantitative manner by mixing peptide mixtures of CRT treated differently e.g., 1 mM Ca^{2+} treated CRT and CRT depleted for Ca^{2+}. Consequently, subtle differences could have been determined. The given cross-linker would restrict the quantitative comparison to two samples at a time, meaning a set of experiments altogether should be performed. This could be circumvented using an isotopically labelled cross-linker available in multiple forms, however, the corresponding mass spectra would be significantly more complicated and difficult to assign. Furthermore, depending on the isotopic cross-linker, there could be quantitational problems in the spectra due to overlapping envelopes even at relatively low mass values.

4. Conclusion

The model of CRT was, for a large part, compatible with the present analysis. The loop structure between Lys24 and Lys38 needs to be remodelled and the residues in the first part of the P-domain need to have reversed orientation. In addition, the first part of the C-domain (up to Lys343) of the molecule was located to run along the surface of the core region opposite the P-domain. The remainder of the C-domain is strongly acidic and may not be in close contact with the core region, as proteolytic cleavage fairly easily reduces CRT to a truncated form by cleavage at Lys334 [35]. The distance measurements by the cross-links found in the C-terminal region are in agreement with the C-terminal at least partially adopting an alpha-helical conformation. The observation that almost all linkages were observed in four independent experiments showed that the method is reproducible, but did not reveal any major changes in conformation upon addition and removal of divalent cations. The combined use of isotopes of BS^3 and ProteinXXX software greatly facilitated the cross-linking analysis. The former because peptides containing the cross-link can be extracted from the large amount of data with relative ease, and the latter because all possible combinations of cross-linker and peptides can be extracted and evaluated in a simplified manner.

Finally, the inclusion of cross-linking data in homology modelling may provide a possibility for directly assessing the quality of the homology models and/or provide distance constraints which may be included in the homology modelling process.

Acknowledgements

Novo Scholarship Programme is gratefully acknowledged for a scholarship to Tina Nielsen and Novo Fonden for a grant to Peter Højrup.

References

[1] K.L. Bennett, T. Matthiesen, P. Roepstorff, Methods Mol. Biol. 146 (2000) 113.
[2] M.M. Young, N. Tang, J.C. Hempel, C.M. Oshiro, E.W. Taylor, I.D. Kuntz, B.W. Gibson, G. Dollinger, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 5802.
[3] J. Rappsilber, S. Siniossolou, E.C. Hurt, M. Mann, Anal. Chem. 72 (2000) 267.
[4] G.T. Hermanson, Bioconjugate Techniques, San Diego, CA, Academic Press, 1996.
[5] G.B. Hurst, T.K. Lankford, S.J. Kennel, J. Am. Soc., Mass Spectrom. 15 (2004) 832.
[6] S.C. Alley, F.T. Ishmael, A.D. Jones, S.J. Benkovic, J. Am. Chem. Soc. 122 (2000) 6126.
[7] E.V. Petrochenko, V.K. Olkhovik, C.H. Borchers, Mol. Cell. Proteomics 4 (2005) 1167.
[8] D.R. Müller, P. Schindler, H. Towbin, U. Wirth, H. Voshol, S. Hoving, M.O. Steinmetz, Anal. Chem. 73 (2001) 1927.
[9] C.J. Collins, B. Schilling, M. Young, G. Dollinger, R.K. Guy, Bioorg. Med. Chem. Lett. 13 (2003) 4023.
[10] E.M. Kosower, N.S. Kosower, Methods Enzymol. 251 (1995) 133.
[11] K.L. Bennett, M. Kussmann, P. Björk, M. Godzwon, M. Mikkelsen, P. Sørensen, P. Roepstorff, Protein Sci. 9 (2000) 1503.
[12] J.J. Peterson, M.M. Young, L.J. Takemoto, Mol. Vision 10 (2004) 857.
[13] X. Tang, G.R. Munske, W.F. Siems, J.E. Bruce, Anal. Chem. 77 (2005) 311.
[14] A. Sinz, Mass Spectrom. Rev. 25 (2006) 663.
[15] B. Schilling, R.H. Row, B.W. Gibson, X. Guo, M.M. Young, J. Am. Soc., Mass Spectrom. 14 (2003) 834.

[16] G.H. Kruppa, J. Schoeniger, M.M. Young, Rapid Commun. Mass Spectrom. 17 (2003) 155.
[17] J. Klose, N. Wendt, S. Kubald, E. Krause, K. Fechner, M. Beyermann, M. Bienert, R. Rudolph, S. Rothemund, Protein Sci. 13 (2004) 2470.
[18] V. Schnaible, S. Wefing, A. Resemann, D. Suckau, A. Bücker, S. Wolf-Kümmeth, D. Hoffmann, Anal. Chem. 74 (2002) 4980.
[19] C. Hammond, I. Braakman, A. Helenius, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 913.
[20] J.D. Schrag, J.J. Bergeron, Y. Li, S. Borisova, M. Hahn, D.Y. Thomas, M. Cygler, Mol. Cell. 8 (2001) 633.
[21] L. Ellgaard, R. Riek, T. Herrmann, P. Guntert, D. Braun, A. Helenius, K. Wuthrich, Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 3133.
[22] J.C. Wu, Z.Q. Liang, Z.H. Qin, Acta Biochim. Biophys. Sin. 38 (2006) 219.
[23] P. Gelebart, M. Opas, M. Michalak, Int. J. Biochem. Cell. Biol. 37 (2005) 260.
[24] S. High, F.J. Lecomte, S.J. Russell, B.M. Abell, J.D. Oliver, FEBS Lett. 476 (2000) 38.
[25] E.S. Trombetta, A.J. Parodi, Ann. Rev. Cell. Dev. Biol. 19 (2003) 649.
[26] J. Roth, M. Ziak, C. Zuber, Biochimie 85 (2003) 287.
[27] G. Houen, C. Koch, Acta Chem. Scand. 48 (1994) 905.
[28] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T.J. Gibson, D.G. Higgins, J.D. Thompson, NAR 31 (2003) 3497.
[29] A. Sali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779.
[30] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornon, J. Appl. Cryst. 26 (1993) 283, accessed via the ADIT validation server at http://deposit.pdb.org/validate/.
[31] M. Kapoor, H. Srinivas, E. Kandiah, E. Gemma, L. Ellgaard, S. Oscarson, A. Helenius, A. Surolia, J. Biol. Chem. 278 (2003) 6194.
[32] P. Højrup, P. Roepstorff, G. Houen, Eur. J. Biochem. 268 (2001) 2558.
[33] C. Wiuff, G. Houen, Acta Chem. Scand. 50 (1996) 788.
[34] Z. Li, W.F. Stafford, M. Bouvier, Biochemistry 40 (2001) 11193.
[35] C.S. Jørgensen, C. Trandum, N. Larsen, L.R. Ryder, M. Gajhede, L.K. Skov, P. Højrup, V. Barkholt, G. Houen, Prot. Pept. Lett. 12 (2005) 687.
[36] W.L. DeLano, The Pymol Molecular Graphics system, DeLano Scientific LLC, San Carlos, CA, USA, 2002, http://www.pymol.org.