doi:10.1016/S0022-2936(02)00812-4 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (2002) 322, 773–784

JMB

NMR Structures of 36 and 73-residue Fragments of the Calreticulin P-domain

Lars Ellgaard[1], Pascal Bettendorff[2], Daniel Braun[2], Torsten Herrmann[2] Francesco Fiorito[2], Ilian Jelesarov[3], Peter Güntert[2], Ari Helenius[1] and Kurt Wüthrich[2]*

[1]*Institut für Biochemie Eidgenössische Technische Hochschule Zürich, CH-8093 Zurich, Switzerland*

[2]*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Zürich CH-8093 Zurich, Switzerland*

[3]*Department of Biochemistry University of Zürich, CH-8057 Zurich, Switzerland*

**Corresponding author*

Calreticulin (CRT) is an abundant, soluble molecular chaperone of the endoplasmic reticulum. Similar to its membrane-bound homolog calnexin (CNX), it is a lectin that promotes the folding of proteins carrying N-linked glycans. Both proteins cooperate with an associated co-chaperone, the thiol-disulfide oxidoreductase ERp57. This enzyme catalyzes the formation of disulfide bonds in CNX and CRT-bound glycoprotein substrates. Previously, we solved the NMR structure of the central proline-rich P-domain of CRT comprising residues 189–288. This structure shows an extended hairpin topology, with three short anti-parallel β-sheets, three small hydrophobic clusters, and one helical turn at the tip of the hairpin. We further demonstrated that the residues 225–251 at the tip of the CRT P-domain are involved in direct contacts with ERp57. Here, we show that the CRT P-domain fragment CRT(221–256) constitutes an autonomous folding unit, and has a structure highly similar to that of the corresponding region in CRT(189–288). Of the 36 residues present in CRT(221–256), 32 form a well-structured core, making this fragment one of the smallest known natural sequences to form a stable non-helical fold in the absence of disulfide bonds or tightly bound metal ions. CRT(221–256) comprises all the residues of the intact P-domain that were shown to interact with ERp57. Isothermal titration microcalorimetry (ITC) now showed affinity of this fragment for ERp57 similar to that of the intact P-domain, demonstrating that CRT(221–256) may be used as a low molecular mass mimic of CRT for further investigations of the interaction with ERp57. We also solved the NMR structure of the 73-residue fragment CRT(189–261), in which the tip of the hairpin and the first β-sheet are well structured, but the residues 189–213 are disordered, presumably due to lack of stabilizing interactions across the hairpin.

Keywords: autonomous folding unit; calreticulin; endoplasmic reticulum; ERp57; NMR structure

Introduction

In the endoplasmic reticulum (ER), two homologous lectin chaperones, calnexin (CNX) and calreticulin (CRT), assist the folding and quality control of proteins carrying N-linked glycans.[1] Whereas CNX is membrane-bound, CRT is a soluble lumenal protein. Both proteins interact specifically with glycoproteins carrying monoglucosylated ($Glc_1Man_{7-9}GlcNAc_2$) trimming intermediates of the triglucosylated core glycan ($Glc_3Man_{7-9}GlcNAc_2$).[2–4] Release from CNX and CRT is ensured by glucosidase II, which removes the remaining glucose from the glycan. Once the glycoprotein has

Present address: P. Güntert, RIKEN Genomic Sciences Center, W505, 1-7-22 Suehiro, Tsurumi, Yokohama 230-0045, Japan.

Abbreviations used: CNX, calnexin; CRT, calreticulin; ER, endoplasmic reticulum; ITC, isothermal titration microcalorimetry; NOE, nuclear Overhauser enhancement; NOESY, NOE spectroscopy; NTA, nitrilotriacetic acid.

E-mail address of the corresponding author: wuthrich@mol.biol.ethz.ch

0022-2836/02/$ - see front matter

adopted its native structure upon release, it is free to leave the ER and proceed along the secretory pathway. Alternatively, if it is still not correctly folded, the glycoprotein is recognized by the UDP-Glc:glycoprotein glucosyltransferase, which thus serves as a folding sensor.[5] By re-adding a glucose unit to the glycan, this enzyme promotes renewed association with CNX and CRT.

Both chaperones also cooperate with the thiol-disulfide oxidoreductase ERp57, which is a close homolog of the protein disulfide isomerase (PDI).[6,7] Like PDI, ERp57 contains four thioredoxin-like domains with active site –CXXC– sequence motifs located in the N and C-terminal domains. *In vivo*, ERp57 promotes disulfide bond formation in glycoprotein substrates bound by CNX and CRT through the formation of transient intermolecular disulfide bonds.[8] Overall, the CNX/CRT chaperone system increases folding efficiency, and prevents aggregation and premature ER exit of newly synthesized glycoproteins.

Both CNX and CRT contain a central proline-rich region, the "P-domain". It consists of two types of sequence repeats with 17 residues (type 1) and 14 residues (type 2), respectively. CNX contains four repeats of each type arranged in a 11112222 fashion, whereas CRT contains three repeats of each type in a 111222 arrangement (Figure 1(a)). The three-dimensional structures of the CNX and CRT P-domains are highly similar, except for the presence of the additional pair of sequence repeats in CNX.[9,10] The CRT P-domain comprising residues 189–288, CRT(189–288), forms an extended hairpin fold with the N and C termini in close proximity (Figure 1(b)).[10] The structure is stabilized by three short antiparallel β-sheets and three small hydrophobic clusters, each involving two tryptophyl rings packed against the aliphatic side-chains of a prolyl and a lysyl residue. The presence of the three equidistantly spaced β-sheets and the three hydrophobic clusters clearly reflects the threefold repetition of sequence repeats in CRT(189–288) (Figure 1).

Recently, the crystal structure of the CNX ectodomain, which includes the P-domain, was solved.[9] It is characterized by a compact lectin domain showing a β-sandwich structure, from which the elongated P-domain extends. Based on the close sequence similarity between the two proteins, the global three-dimensional structure of

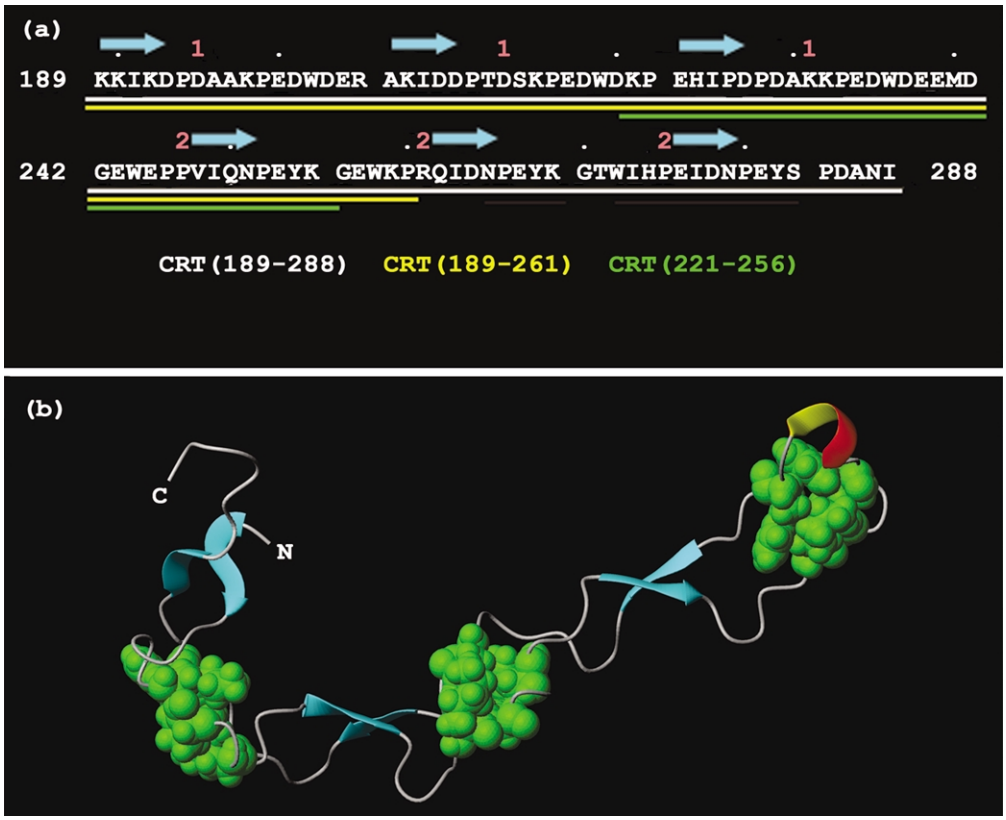

Figure 1. (a) Chemical structures of three CRT P-domain constructs. The amino acid sequence of rat calreticulin (GenBank accession number CAA55890) is shown for residues 189–288. The colored lines below the sequence indicate lengths and locations of the fragments CRT(221–256) (green), CRT(189–261) (yellow) and CRT(189–288) (white). The positions of the β-sheets in the NMR structure of CRT(189–288) are indicated by blue arrows above the sequence. White dots mark every tenth residue, beginning at residue number 190. Spaces between residues indicate boundaries between subsequent sequence repeats, and the red numbers above the sequence identify the type of repeat (see the text). (b) Three-dimensional structure of CRT(189–288). The ribbon drawing shows the β-sheets in cyan and an α-helical turn in red. The three hydrophobic clusters are shown as green all-heavy-atom space-filling models.

CRT can be assumed to be similar to that of CNX. In the linear sequence of the two proteins, the P-domain is inserted in-between the two peripheral regions, which are in close proximity in the three-dimensional structure and form the β-sandwich and a structural Ca^{2+}-binding site.

The role of the P-domain in glycoprotein folding has become clearer with our recent finding that the CRT P-domain interacts with the co-chaperone ERp57.[11] Specifically, we were able to show that the interaction with ERp57 occurs through residues 225–251 at the tip of the P-domain structure.[11] Thus, in the binary CRT–ERp57 complex, the lectin domain, the P-domain and ERp57 appear to form a partially solvent-shielded "reaction chamber", where folding of the bound glycoprotein could take place while access for other folding intermediates and chaperones is restricted. In addition, the positioning of ERp57 at the tip of the P-domain might facilitate the interaction with cysteine residues in the substrate glycoprotein.

Here, we describe the structure determination by solution NMR spectroscopy of two polypeptide fragments from the CRT P-domain, CRT(221–256) and CRT(189–261), consisting of 36 and 73 residues, respectively. The results indicate that CRT(221–256) is a promising candidate for use in further studies into the nature of the interaction between CRT and ERp57, since it comprises the region of CRT that is involved in the direct contacts with ERp57,[11] and is demonstrated to interact with ERp57 with comparable affinity as the intact CRT P-domain.

Results

Design, expression and physical–chemical characterization of the CRT P-domain fragments CRT(221–256) and CRT(189–261)

To identify potential independently folding fragments of the CRT P-domain, limited proteolysis experiments were carried out on a CRT construct comprising residues 189–300. A subtilisin-resistant fragment comprising residues 189–261 was identified by mass spectroscopy and N-terminal sequencing (K. Ng, J. Peterson, W. Weis & A.H., unpublished data). As shown in Figure 1, this fragment encompasses the three type 1 repeats and approximately the first one and a half type 2 repeats. In the structure of CRT(189–288),[10] the residues 189–261 form two hydrophobic clusters and one β-sheet, but C-terminally this fragment lacks the residues involved in the formation of the third hydrophobic cluster as well as the two other β-sheets. To identify potential smaller independently folding segments in CRT(189–288), we used the RAFT (rapid autonomous fragment test)[12] algorithm for prediction of continuous autonomous folding units (for details, see Materials and Methods). Thereby a fragment comprising residues 221–256 was identified (Kael F. Fischer, personal communication). This fragment is centered around the hairpin tip and contains exclusively residues involved in the formation of the hydrophobic cluster and the β-sheet nearest to the hairpin tip.

The two fragments CRT(221–256) and CRT(189–261) were cloned into suitable *Escherichia coli* expression vectors (see Materials and Methods). Both constructs expressed at high levels, and were purified in a two-step procedure involving affinity purification on a Ni-NTA column followed by anion-exchange chromatography. For the NMR structure determination, CRT(189–261) was labeled with ^{13}C and ^{15}N, whereas CRT(221–256) was not labeled.

Thermal denaturation curves of CRT(221–256) and CRT(189–261) were recorded with circular dichroism (CD) spectroscopy at 222 nm in the temperature range 5–80 °C, and compared to that of the full-length P-domain, CRT(189–288). Complete reversibility of denaturation was observed for all three proteins upon returning to the initial temperature of 5 °C, with t_m values of 32 °C for the two fragments, and 39 °C for CRT(189–288). The thermal denaturation temperature of CRT(189–288) is only slightly lower than that of $t_m = 42.5$ °C obtained for full-length CRT.[13,14] The low thermal stability of full-length CRT and its intrinsic flexibility have been suggested to be of functional importance for the protein with respect to its role as a molecular chaperone.[13,14] Overall, the data showed that although CRT(221–256) and CRT(189–261) exhibit low thermal stability, the folded form of both fragments is highly populated in solution at low temperature.

Structural integrity of the two P-domain fragments is also indicated by 1D ^1H NMR spectra, which show well-dispersed signals that indicate a folded conformation (Figure 2).[15,16] In particular, based on the resonance assignments of CRT(189–288),[10] we assigned the ring current-shifted resonances of Lys232 between 0 and 1 ppm in the spectra of both CRT(221–256) and CRT(189–261), which also contain the characteristic indole $H^{\varepsilon 1}$ lines of Trp236 at 10.6 ppm and Trp244 $H^{\varepsilon 1}$ at 9.9 ppm.[16] In addition, the resonance line of H^{α} of Gln250 observed near the water line in the spectrum of CRT(189–261) indicates that a β-sheet involving Gln250 is formed in this construct.[15]

Resonance assignments, collection of conformational constraints and NMR structure determination

The NMR solution structures of CRT(221–256) and CRT(189–261) were solved completely independently. The sequential assignment of CRT(189–261) was based on uniform ^{13}C,^{15}N-labeling and standard triple resonance and nuclear Overhauser enhancement spectroscopy (NOESY) experiments.[17–19] All 61 backbone amide resonances expected in a 2D [^{15}N,^1H] correlated spectroscopy (COSY) spectrum were identified and

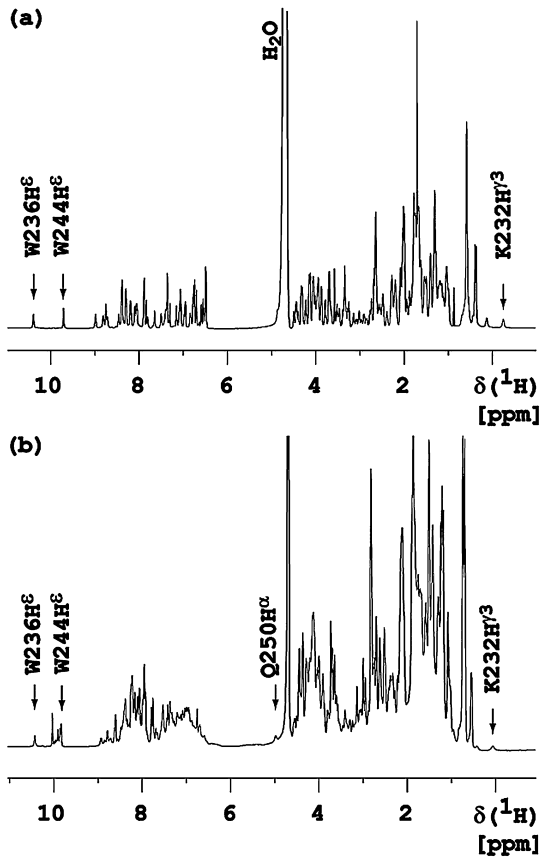

Figure 2. One-dimensional 800 MHz ^{1}H NMR spectra of CRT(221–256) (a) and CRT(189–261) (b). The spectra were recorded at 7 °C for CRT(221–256) and at 20 °C for CRT(189–261), using the solvent conditions given in Materials and Methods. Resonance assignments are given for selected resonances (see the text).

assigned from a HNCACB spectrum[17] in combination with sequential NOE connectivities.[18,19] All 12 Xxx–Pro dipeptide segments were identified based on sequential $d_{\alpha\delta}$ and $d_{\beta\delta}$ NOE connectivities.[15] The assignments for the backbone protons and the non-labile protons of the amino acid side-chains are complete, except for $H^{\varepsilon 3}$ and $H^{\zeta 3}$ of Trp202 and Trp219, and $H^{\varepsilon 1}$ of His224.

For CRT(221–256), the assignment was based on sequential NOEs, using standard homonuclear procedures.[15] All 31 intraresidual H^N–H^α resonances expected in the homonuclear 2D double quantum filtered (DQF)-COSY spectrum were identified, and the spin system identifications were obtained from a 2D [^{1}H,^{1}H] total correlated spectroscopy (TOCSY) spectrum. Sequential $d_{\alpha N}$ connectivities were observed for all non-proline residues, and sequential $d_{\alpha\delta}$ NOE connectivities were identified for all seven Xxx–Pro dipeptide segments. The assignments are complete except for $H^{\varepsilon 1}$ of His224.

A total of 1493 NOE cross-peaks for CRT(221–256) and 3777 NOE cross-peaks for CRT(189–261) were used along with the chemical shift lists derived from the sequence-specific assignments as input for the programs CANDID[20] and DYANA[21] (see Materials and Methods). In addition, scalar coupling constants were extracted from a 2D [^{1}H,^{1}H]-exclusive COSY (ECOSY) spectrum to obtain 102 dihedral angle constraints. The calculations yielded assignments of 448 and 1297 meaningful NOE upper distance limits for CRT(221–256) and CRT(189–261), respectively, as well as the NMR structures of both fragments (see below and Table 1).

Even before inspection of the final protein structures, patterns of long-range NOEs in CRT(221–256) and CRT(189–261) indicated highly similar overall folds. This is illustrated in Figure 3 by spectral regions containing resonances of Trp236 and Trp244, which are involved in the formation of the hydrophobic cluster near the tip of the hairpin in CRT(189–288). Several long-range NOEs are observed for both CRT(221–256) and CRT(189–261), including interresidual NOEs between H^N and $H^{\varepsilon 1}$ of the two tryptophyl residues, as well as NOEs between Trp236 and Trp244 and side-chain protons of Lys232 and Pro233. More generally, as shown in Figure 4, the region of the diagonal plots comprising the residues of the smaller fragment contains the typical hairpin pattern previously observed for CRT(189–288).[16] The diagonal plot of CRT(189–261) contains no long-range NOEs between residues 189 and 213, indicating the presence of a disordered tail. This result was corroborated by steady-state ^{15}N{^{1}H}-NOE values that are compatible with increased mobility in this segment (data not shown).

For CRT(221–256), we exercised special care to characterize possible conformational polymorphisms. Figure 4(a) shows that except for

Table 1. Input for the structure calculation and characterization of the energy-minimized NMR structures of CRT(221–256) and CRT(189–261)

Quantity[a]	CRT(221–256)	CRT(189–261)
NOE upper distance limits	448	1297
Dihedral angle constraints	102	202
Residual target function (Å^2)	0.34 ± 0.08	1.12 ± 0.34
Residual distance constraint violations		
Number ≥ 0.1 Å	2 ± 1	6 ± 2
Maximum (Å)	0.12 ± 0.07	0.13 ± 0.10
Residual dihedral angle constraint violations		
Number ≥ 2.5 deg.	1 ± 1	0 ± 1
Maximum (deg.)	2.81 ± 0.05	2.30 ± 1.07
AMBER energies (kcal/mol)		
Total	−658 ± 61	−1607 ± 232
Van der Waals	−68 ± 5	−107 ± 13
Electrostatic	−927 ± 60	−2184 ± 220
r.m.s. deviations from ideal geometry		
Bond lengths (Å)	0.0071 ± 0.0002	0.0076 ± 0.0003
Bond angles (deg.)	1.892 ± 0.046	2.079 ± 0.062
r.m.s. deviation from the mean coordinates (Å)		
N, C^α, C′ (223–254)	0.43 ± 0.13	0.52 ± 0.19
All heavy atoms (223–254)	0.98 ± 0.13	1.01 ± 0.19
N, C^α, C′ (214–258)	–	0.89 ± 0.31
All heavy atoms (214–258)	–	1.30 ± 0.31

[a] Except for the top two entries, the data characterize the group of 20 conformers that is used to represent each NMR structure; the mean value and the standard deviation are given.

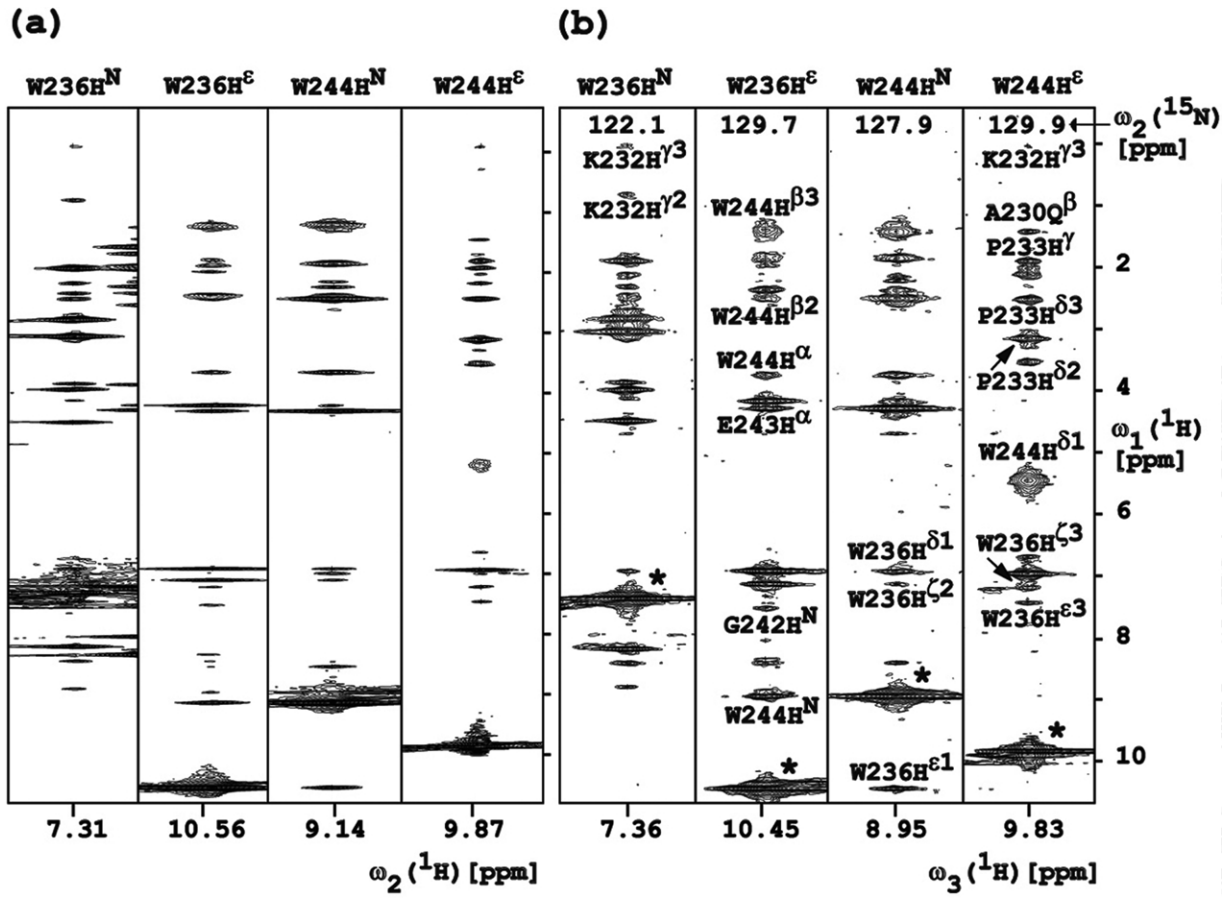

Figure 3. Long-range NOEs observed in CRT(221–256) and CRT(189–261). (a) Contour plot from a 2D [^{1}H,^{1}H]-NOESY spectrum of CRT(221–256). The four spectral regions are centered about the H^N and H^ε chemical shifts of the residues Trp236 and Trp244, as indicated at the top. (b) [ω_1(^{1}H), ω_2(^{1}H)] strips from a 3D ^{15}N-resolved [^{1}H,^{1}H]-NOESY spectrum of CRT(189–261). The strips were taken at the backbone ^{15}N and indole $^{15}N^\varepsilon$ chemical shifts of Trp236 and Trp244, as indicated at the top, and are centered about the H^N or H^ε chemical shifts along ω_3(^{1}H). They correspond to the four strips in (a), although the $^1H^N$ chemical shifts are somewhat different due to the different temperatures at which the spectra were recorded, i.e. 7 °C and 20 °C, respectively. Diagonal peaks are labeled with asterisks, and resonance assignments by the one-letter amino acid symbols and the sequence numbers are given for those long-range NOEs that are also observed in (a). The resonance assignment of the ring current-shifted Trp244 $H^{\delta1}$ is also shown (see the text).

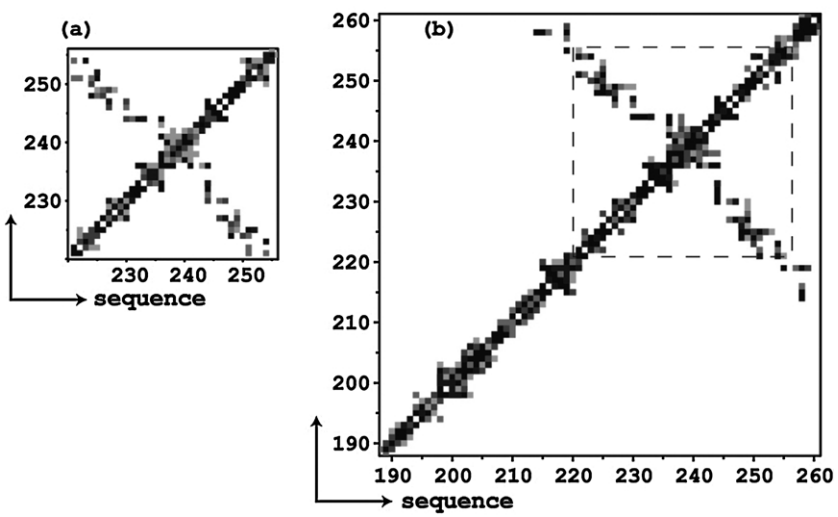

Figure 4. Diagonal plots of the NOE upper distance constraints identified in CRT(221–256) (a) and CRT(189–261) (b). The sequence numbering is shown on both axes. The presence of a distance constraint between a pair of residues is indicated by a square. Increasing darkness of the squares indicates an increasing number of NOE constraints between the two residues, with black squares representing five or more NOEs. No distinction is made between NOEs involving backbone or side-chain hydrogen atoms. In (b), the region corresponding to CRT(221–256) is indicated by broken lines.

the two C-terminal residues, the entire sequence is involved in long-range NOEs. Indications for the existence of minor conformations were limited to two weak unassigned peaks in the [^{1}H,^{1}H]-COSY fingerprint,[15] one of which was located next to the signal of the C-terminal Gly256. We concluded that at 7 °C there is a major conformation of CRT(221–256) populated to >95%. Spectra recorded at 20 °C showed no significant line broadening (data not shown), indicating that there is no conformational exchange on the intermediate exchange time-scale in the NMR sample of CRT(221–256) used for the structure determination.

The NMR solution structures of CRT(221–256) and CRT(189–261)

CRT(221–256) and CRT(189–261) exhibit an extended hairpin fold similar to that of CRT(189–288) (Figures 5 and 6). Superposition of the backbone heavy atoms of residues 223–254 of the 20 conformers of CRT(221–256) and CRT(189–261) (Figure 7) shows an r.m.s. deviation of 1.7 Å, illustrating the close structural similarity between the two fragments. Both fragments contain a helical turn at the tip of the hairpin and a short anti-parallel β-sheet comprising residues 224–226 and 248–250, which are also present in the structure of CRT(189–288) (Figure 1(b)).[10,16] Similarly, a hydrophobic cluster near the tip of the hairpin involving the rings of two conserved tryptophyl residues, Trp244 and Trp236, Lys232 and Pro233 is a conserved feature from the structure of CRT(189–288) (Figures 1(b), 5 and 6). In CRT(189–261), a second hydrophobic cluster involving Lys215, Pro216, Trp219 and Trp258 can be seen but is less well defined (Figure 6). The lack of long-range stabilizing interactions for residues 189–213 in CRT(189–261) (Figure 4(b)) goes along with the presence of an N-terminal disordered tail (Figure 6). In contrast, CRT(221–256) is well-ordered throughout the whole structure, with an r.m.s. deviation value of 0.43(±0.13) Å for the backbone heavy atoms N, C^{α} and C′ of residues 223–254, and of 0.98(±0.13) Å when calculated for all heavy atoms in this segment (Figure 5(a) and Table 1).

Discussion

The present study shows that the unusual hairpin-type fold of the 100-residue CRT P-domain is maintained in two smaller fragments of this polypeptide chain, containing 73 and 36 residues, respectively. The fact that the larger subdomain, CRT(189–261), contains an N-terminal disordered tail of residues 189–213 (Figures 4(b) and 6), further illustrates that the formation of this fold-type depends critically on inter-strand contacts in the hairpin, since the partner strand segment for residues 189–213 is missing in this construct. CRT(221–256) forms a stable fold with a core of 32 residues, which retains the structural features

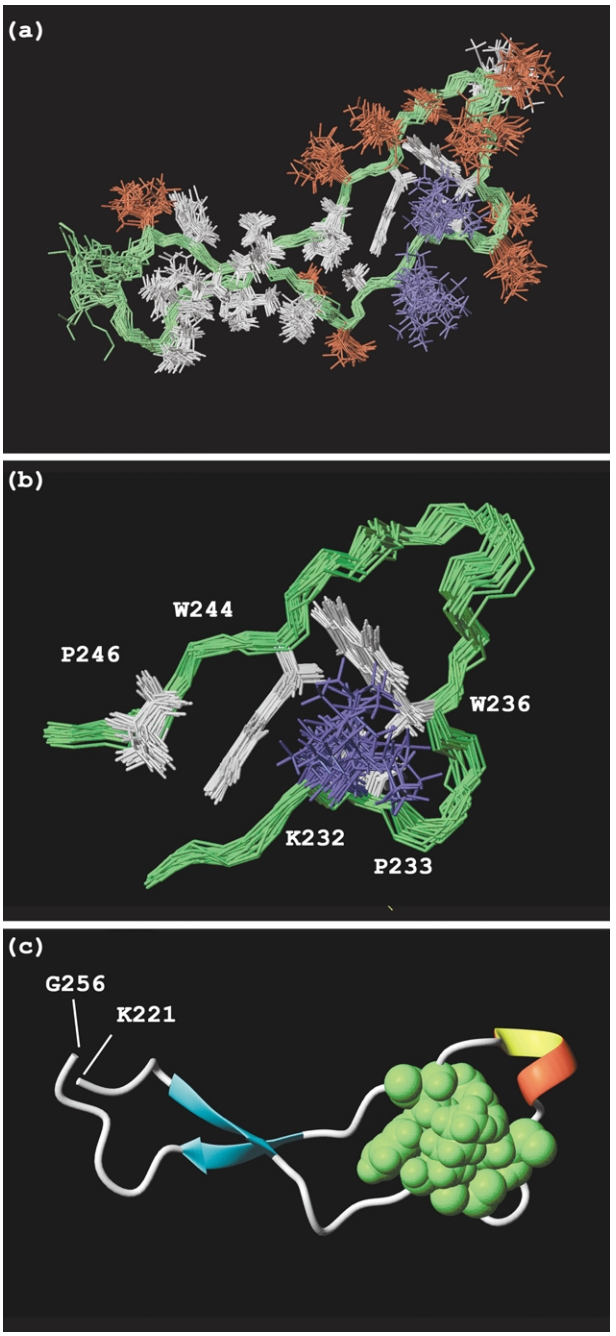

Figure 5. (a) and (b) Bundles of the 20 energy-minimized conformers used to represent the NMR structure of CRT(221–256) after superposition for best fit of the backbone atoms N, C^{α} and C′ of the residues 223–254. (a) All-heavy-atom presentation of the complete structure. The backbone is colored green, positively charged residues are blue, negatively charged residues are red, and hydrophobic and polar residues are white. (b) Close-up view showing the side-chain arrangement of the residues Lys232, Pro233, Trp236, Trp244 and Pro246 in CRT(221–256), which are all affected by ring-current shifts due to proximity to the indole rings (Table 2). (c) Ribbon drawing of one of the 20 CRT(221–256) conformers shown in (a). The β-sheet is cyan, the α-helical turn is red, and the residues Lys232, Pro233, Trp236 and Trp244 of the hydrophobic cluster are shown in green as all-heavy-atom space-filling models.

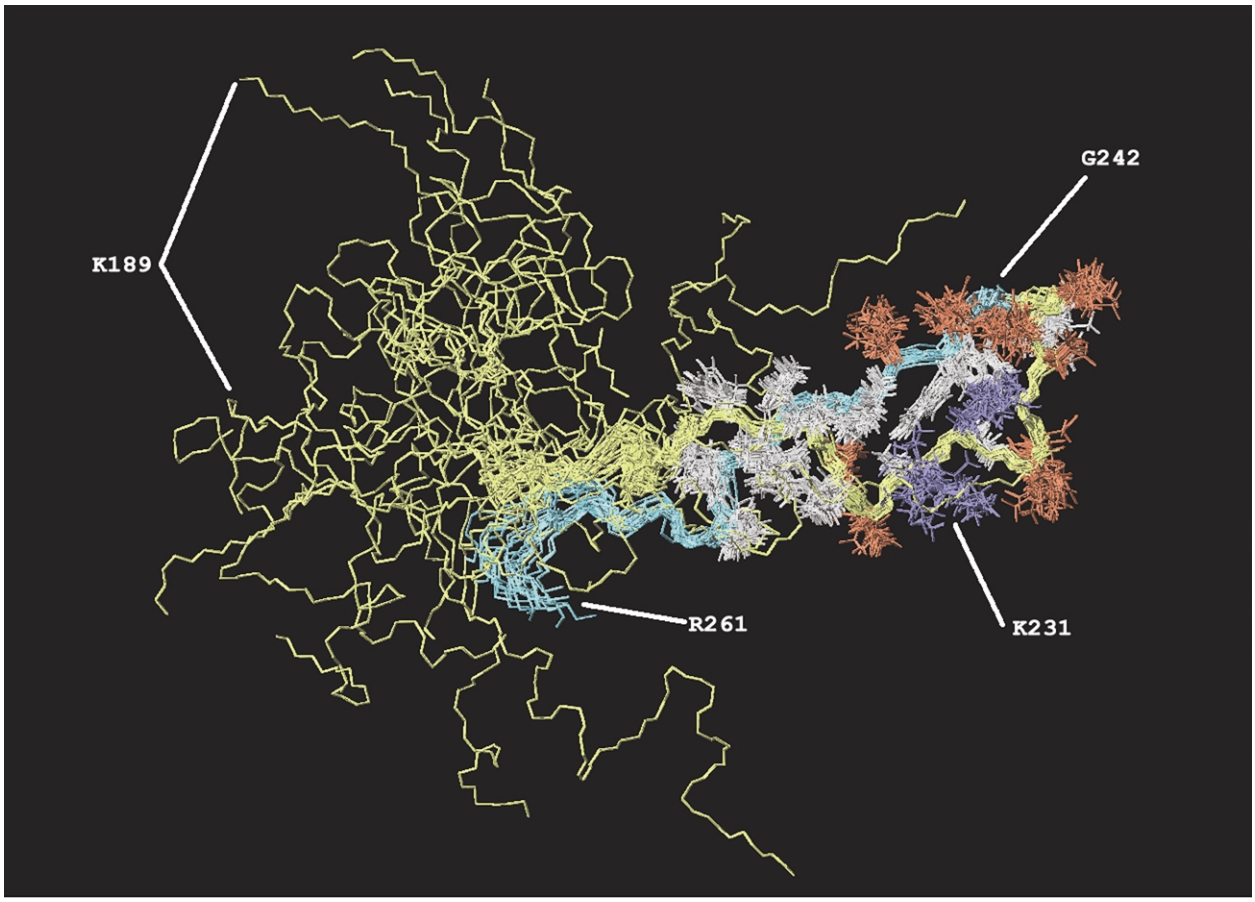

Figure 6. Bundle of the 20 energy-minimized conformers used to represent the NMR structure of CRT(189–261). Superposition of the polypeptide chain for best fit of the backbone atoms N, C^{α} and C' has been performed for the residues 214–258. The backbone is yellow for residues 189–241 and cyan for residues 242–261, and the side-chains are color-coded as in Figure 5(a). The N-terminal chain end is indicated for two conformers, and certain other positions are identified for the entire bundle.

of the corresponding polypeptide segment in CRT(189–288).[10] This 36-residue polypeptide is thus of keen interest, on the one hand because it represents one of the smallest known polypeptides forming a stable non-helical structure in the absence of disulfide bonds or metal ions, and on the other hand because it comprises all residues involved in direct interactions between the CRT P-domain and ERp57.[11]

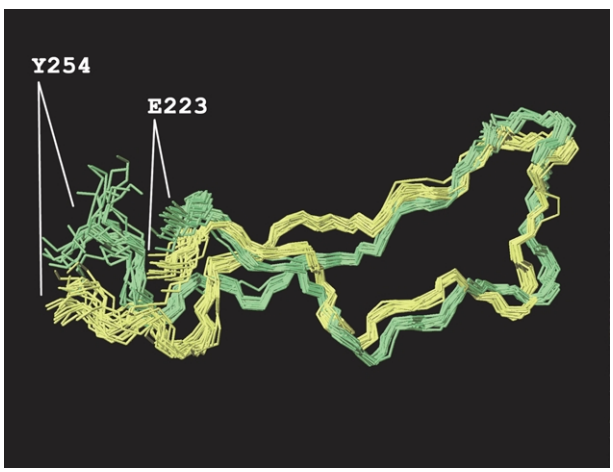

Figure 7. Superposition of the residues 223–254 of CRT(221–256) (green) and CRT(189–261) (yellow) for best fit of the backbone heavy atoms, with a mean r.m.s. deviation of 1.7 Å. For each of the two structures a set of 20 conformers is shown, which had previously been fitted individually to the respective mean coordinates. The two chain ends are indicated for both structures.

In NOE-based NMR structure determinations with marginally stable molecular systems, for example, residual non-random conformations of polypeptides in denaturing solvents or truncated fragments of stably folded domains, one may identify combinations of local features that are not simultaneously present in the same molecules, or detect a global fold adopted only by a fraction of the ensemble of molecules.[22] It is therefore of critical importance to further characterize the NMR structure on the basis of parameters with different ensemble-averaging than the r^{-6}-dependence of the NOEs.[15] In CRT(221–256), the presence of the indole rings of Trp236 and Trp244 makes an analysis of ring current effects on the chemical shifts of nearby protons an obvious choice.[23,24] Using this approach, the chemical shift predictions based on the NMR structure of CRT(221–256) (Figure 5(b)) and the experimental data were found to be in good qualitative agreement (Table 2). For several resonances the shifts predicted by the ring current

calculations are even larger than the experimental values. Based on previous experience,[23,24] the ring current calculations show that the NMR structure of CRT(221–256) is simultaneously adopted by a high population, near 100%, of the protein molecules in solution at 7 °C.

There is only a small number of naturally occuring protein subdomains or designed mini-protein motifs with about 40 residues that have been found to be stably folded in the absence of disulfide bonds or coordinatively bound metal ions.[25] These include different WW domains (34–37 residues),[26] the peripheral subunit-binding domain from the pyruvate dehydrogenase multienzyme complex (43 residues),[27–29] and the villin headpiece subdomain, HP-35 (35 residues).[30,31] The reduced complexity of experimental and computational data when compared to larger domains makes these small structures attractive tools for in-depth studies of sequence–structure relations and protein folding.[32] In particular, the identification of CRT(221–256) as an independently folding subdomain may help to gain new insights into the physical–chemical foundations for the unusual structure of the CRT P-domain.

An extended hairpin fold containing three short anti-parallel β-sheets has also been found in the cysteine-rich (CR) domain of the molecular chaperone DnaJ from *E. coli*.[33] However, in the CR domain of DnaJ the hydrophobic clusters observed in the CNX and CRT P-domains are substituted by clusters containing one zinc ion and one –CXXC-XGXG– sequence motif on each of the two hairpin strands. In the absence of zinc the CR domain unfolds, indicating a structural role of the zinc ions.[33] In view of the presently described structure of CRT(221–256) it would now seem intriguing to investigate the existence of independently folding subdomains containing one or two Zn^{2+} clusters of the DnaJ CR domain.

The current study also bears on the process by which CNX and CRT might fold *in vivo*. As documented by the structure of the CNX ectodomain,[9] a globular lectin domain is formed by sequences flanking the P-domain on both the N and C-terminal sides. Since the β-sandwich and the Ca^{2+}-binding site involve residues on the C-terminal side of the P-domain, the formation of the entire lectin domain probably does not occur until after complete folding of the P-domain. As the nascent chain is being synthesized, the N-terminal strand of the P-domain hairpin is likely to remain unfolded until the residues of the opposite strand of the hairpin are synthesized (see also Figure 6). This then quite naturally leads to the hypothesis that P-domain folding proceeds from the hydrophobic cluster nearest to the hairpin tip in the direction of the C terminus, sequentially forming the three β-sheets and the two remaining hydrophobic clusters, much like the closure of a zipper. Presumably this P-domain folding mode would then also lead to the proper positioning of the polypeptide chain for the completion of the β-sandwich and the Ca^{2+}-binding site.

Finally, the previous mapping of the binding site for the CRT and CNX-associated co-chaperone ERp57 to residues 225–251 of the CRT P-domain[11] indicated that CRT(221–256) might not only be a structural subdomain but also a functional subdomain of CRT, capable of interacting independently with ERp57. Indeed, isothermal titration microcalorimetry (ITC) performed at 8 °C detected a measurable heat effect upon mixing of the two proteins. The system reached saturation with increasing molar ratio of CRT(221–256) to ERp57 (Figure 8), and the shape of the resulting titration curve corresponds to a typical binding isotherm with 1:1 stoichiometry. The data could be fitted with a dissociation constant of

Table 2. Ring current shifts in CRT(221–256)

^{1}H atom	$\Delta\delta_{obs}$(ppm)	$\Delta\delta_{rc}$(ppm)
K232H^{N}	−0.83	−0.78
K232H$^{\alpha}$	−2.56	−4.17
K232H$^{\beta2}$	−1.02	−1.06
K232H$^{\beta3}$	−0.95	−0.70
P233H$^{\alpha}$	−0.35	−0.28
P233H$^{\beta2}$	0.05	0.06
P233H$^{\beta3}$	−0.06	−0.05
W244H^{N}	0.82	0.39
W244H$^{\alpha}$	−1.04	−0.27
W244H$^{\beta2}$	−0.88	−0.57
W244H$^{\beta3}$	−1.98	−1.96
P246H$^{\alpha}$	−0.98	−1.68
P246H$^{\beta2}$	−0.27	−0.26
P246H$^{\beta3}$	−0.44	−0.33

Observed chemical shift deviations from the random coil values are compared with the corresponding chemical shift deviations predicted from the ring current shifts calculated to arise from Trp236 and Trp244.

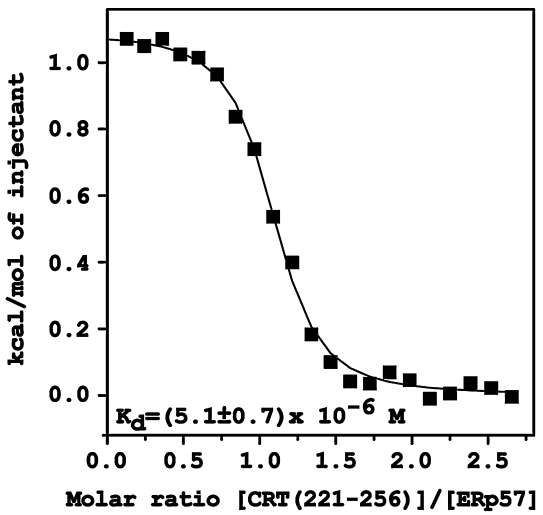

Figure 8. Binding isotherm manifesting the formation of the CRT(221–256)/ERp57 complex measured by ITC at 8 °C. Filled squares represent the integrated heats at each injection after correction for non-specific heat effects and normalization for the molar concentration. The continuous line visualizes a non-linear least-squares fit of the data to a 1:1 binding model defined by the following parameters: $n = 1.05(\pm 0.01)$, $K_d = 5.1(\pm 0.7) \times 10^{-6}$ M ($K_a = 2.0(\pm 0.26) \times 10^5$ M^{-1}).

$5.1(\pm 0.7) \times 10^{-6}$ M, as compared to the value of $9.1(\pm 3.0) \times 10^{-6}$ M previously determined for the interaction of ERp57 with the intact P-domain.[11] It is thus conceivable that CRT(221–256) might become an interesting tool for further biochemical and structural characterization of the CRT/ERp57 chaperone system. For example, mutational analysis of protein–protein interactions and the generation of new antibodies might be performed using the CRT(221–256) construct, or stabilized variants thereof, instead of the entire CRT P-domain. Overall, the structure determination of CRT(221–256) appears to present a platform for novel studies on protein folding as well as on mechanistic aspects of the CRT/ERp57 chaperone system, and possibly even for pharmacological studies targeting these ER chaperones.

Materials and Methods

Rapid autonomous fragment test (RAFT) calculation

As described in detail by Fischer & Marqusee,[12] the RAFT algorithm first identifies all inter-residue contacts present within a protein of known three-dimensional structure. Next, continuous segments are scored based on the number of internal contacts within a given segment as compared to the number of external contacts. Finally, the score is normalized to the length of the segment. For the present calculation, the CRT(189–288) structure was used as input. The shortest fragment identified by the RAFT algorithm within CRT(189–288) comprised residues 221–256, for which a RAFT score of 20.69 contacts/residue was obtained. As discussed by Fischer & Marqusee,[12] a cutoff value of 20 contacts/residue is taken to be indicative of a potential autonomous folding unit, although not all fragments with RAFT scores of ≥ 20 contacts/residue have indeed been found to be structured.[12] The CRT(189–288) structure has been added to the RAFT database†, under the 1hhn PDB entry code.

Construction of expression plasmids

The pGEX-CRTwt vector[34] containing the full-length rat calreticulin cDNA (GenBank accession number X79327) was used as the template in PCR reactions to generate fragments encoding the amino acid residues 189–261 and 221–256 of the protein. For CRT(221–256), the resulting PCR product was cloned into the pT_7H_6UB *Escherichia coli* expression vector.[35] When cloned into this vector, the construct encodes an N-terminal His_6-sequence followed by residues 2–76 of human ubiquitin, a factor X_a (FX_a) cleavage site, and the calreticulin sequence. After FX_a cleavage, a single extra glycine residue remains at the N terminus of the CRT(221–256) construct. For CRT(189–261), the PCR product was cloned into a pRSET A-derived *E. coli* expression vector, as previously described for the CRT(189–288) construct.[10] Both constructs were verified by DNA sequencing.

† http://zebra.berkeley.edu/raft

Protein sample preparation

To produce recombinant unlabeled CRT(221–256), 1000 ml of LB medium containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) was inoculated with 10 ml of overnight culture of *E. coli* BL21(DE3)-pLysS cells that had been freshly transformed with expression plasmid. The culture was grown at 37 °C until reaching an $A_{600\,nm}$ of 0.8, and induction of expression was initiated by the addition of 1 mM isopropyl β-D-galactopyranoside (IPTG). Incubation was continued for three hours post-induction, at which point the cells were harvested by centrifugation and resuspended in 50 ml of buffer A (6 M guanidinium–HCl, 50 mM Tris–HCl (pH 8.0), 10 mM reduced glutathione) containing 100 mM NaCl. Following sonication, the clear supernatant obtained by centrifugation at 20,000*g* for 30 minutes was applied to a Ni^{2+}-charged nitrilotriaceticacid (NTA) metal-chelate affinity column, from which contaminant proteins were eluted during further washing with buffer A containing 1 M NaCl. Applying a linear buffer gradient, the guanidinium–HCl wash buffer was gradually replaced by an aqueous buffer B (500 mM NaCl, 50 mM Tris–HCl (pH 8.0), 2 mM $CaCl_2$). After washing the column with buffer B containing 25 mM imidazole, the fusion protein was eluted from the Ni^{2+}-NTA column with buffer B containing 500 mM imidazole. Removal of the N-terminal fusion tail was achieved by FX_a cleavage, which was typically performed for 24 hours at room temperature using a ratio of 1:75 (w/w) of FX_a to fusion protein. Following cleavage, the protein solution was diluted 20-fold with buffer C (25 mM NaCl, 25 mM Tris–HCl (pH 8.0)) and loaded onto a MonoQ 10/10 anion-exchange column (Pharmacia) before elution with a linear gradient against buffer C containing 500 mM NaCl. Finally, the protein was gel-filtered into 50 mM potassium phosphate buffer (pH 6.5) containing 25 mM NaCl, and concentrated to 5.3 mM using Centricon spin columns (Amicon). The yield of pure CRT(221–256) was 25 mg per liter of LB medium. Recombinant uniformly ^{13}C,^{15}N-labeled CRT(189–261) was prepared as described for CRT(189–288).[16] The correct molecular mass of both proteins was verified by matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry.

Circular dichroism measurements

Thermal denaturation experiments by CD spectroscopy for all three constructs, CRT(221–256), CRT(189–261) and CRT(189–288), were performed on a Jasco J-810 spectropolarimeter. The protein concentration was 30 μM in 5 mM potassium phosphate buffer (pH 6.5) containing 0.5 mM NaCl. The temperature was raised from 5 °C to 80 °C at a rate of 2 deg. C/minute, and the CD signal at 222 nm was monitored at 0.5 deg. C intervals. The response time was 16 seconds and the bandwidth 2 nm. The quartz cuvette cell length was 0.1 cm. In all experiments, 100% of the CD signal intensity was recovered after returning to the initial temperature of 5 °C, showing that the denaturation is reversible for all three proteins.

NMR spectroscopy

The NMR measurements were performed on Bruker DRX600, DRX750 or DRX800 spectrometers equipped with four radio-frequency channels and triple resonance probe heads with shielded *z*-gradient coils. The ^{1}H, ^{15}N

and ^{13}C chemical shifts were calibrated relative to 2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt (DSS).

For the structure determination of CRT(189–261) we used a 3 mM solution of the uniformly ^{13}C,^{15}N-labeled protein in 95% H_2O/5% ^2H_2O containing 10 mM $CaCl_2$ at pH 6.3. The NMR measurements were performed at $t = 20$ °C. For the resonance assignments and collection of conformational constraints the following experiments were recorded: 3D HNCACB,[17] 3D CBCA(CO)NH,[36] 3D *ct*-HCCH-TOCSY with 14 ms mixing time,[37] 3D ^1H-TOCSY-relayed ct-[^{13}C,^1H]-HMQC,[38] 3D combined ^{15}N/^{13}C-resolved [^1H,^1H]-NOESY[39,40] in H_2O, and $^{13}C^R$-resolved [^1H,^1H]-NOESY[39] in H_2O. All NOESY spectra were recorded with a mixing time of $\tau_m = 60$ ms.

For the structure determination of CRT(221–256) we used a 5.3 mM solution of the protein with natural isotope abundance in 90% H_2O/10% ^2H_2O containing 25 mM NaCl and 50 mM potassium phosphate buffer at pH 6.5. The NMR measurements were performed at $t = 7$ °C. Spin system assignment was performed using homonuclear 2D DQF-COSY[41] and 2D TOCSY[15] with a mixing time of $\tau_m = 100$ ms. Coupling constants were derived from a 2D [^1H,^1H]-ECOSY spectrum.[42] For the resonance assignments and the collection of conformational constraints we used a 2D [^1H,^1H]-NOESY spectrum[15] recorded with a mixing time of $\tau_m = 60$ ms.

Collection of conformational constraints and calculation of the three-dimensional structures

For both CRT(221–256) and CRT(189–261), combined NOE cross-peak assignment and 3D protein structure calculation were performed using the programs CANDID[20] and DYANA.[21] The input consisted of the chemical shift lists obtained from the previous sequence-specific resonance assignment, and the listings of cross-peak positions and volumes. In combination, CANDID and DYANA perform automated NOE assignment, automated calibration and violation analysis of NOE upper distance limits, automated generation of upper distance constraints, and calculation of the 3D protein structure represented as a bundle of conformers. Three peak lists for CRT(189–261) and a single peak list for CRT(221–256) were generated by interactive peak picking of the NOESY spectra with the program XEASY.[43] For CRT(189–261), automatic integration of the peak volumes was performed with the program SPSCAN (Ralf Glaser, unpublished). For CRT(221–256), the peaks were manually integrated using the *peakint* routine of the XEASY package. For CRT(221–256), scalar coupling constants were extracted from a 2D [^1H,^1H]-ECOSY spectrum.[42] In each CANDID cycle, these scalar coupling constants were converted in conjunction with the updated list of NOE upper distance constraints into torsion angle constraints by the grid search procedure FOUND.[44]

The iterative CANDID/DYANA procedure[20] comprised seven cycles of NOE assignment and structure calculation. During the first six cycles, CANDID uses ambiguous distance constraints.[45] In the final CANDID cycle, only distance constraints were retained which could be attributed to a single pair of hydrogen atoms.[20] The 20 conformers of the CANDID/DYANA cycle 7 with the lowest final DYANA target function values were energy-minimized in a water shell with the program OPALp,[46,47] using the AMBER force field.[48] The program MOLMOL[49] was used to analyze the resulting 20 energy-minimized conformers and to prepare drawings of the structures.

Ring current shift calculations

Calculations of conformation-dependent chemical shifts[15] due to ring current effects were performed with the program MOLMOL[49] using the Johnson–Bovey algorithm.[23,50] The 20 energy-minimized conformers of CRT(221–256) were used as input for these calculations, and the mean of the 20 values is given in Table 2.

Isothermal titration calorimetry

ITC was performed at 8 °C on an MCS instrument (MicroCal, Inc., Northampton, MA) calibrated either with electrically generated heat pulses or by measurements of the heat of standard chemical reactions. CRT(221–256) and ERp57 prepared as described[11] were both dialyzed into an aliquot of the same batch of 25 mM Tris–HCl buffer (pH 7.0) with 10 mM β-mercaptoethanol. Protein concentrations after dialysis were determined from the absorbance at 280 nm, using molar extinction coefficients calculated by the method of Gill & von Hippel.[51] The cell was loaded with 1.36 ml of 0.23 mM ERp57. The titration protocol consisted of 24 12 μl injections of a 3.02 mM solution of CRT(221–256). Injection duration was ten seconds and equilibration was allowed for five minutes between injections. The stirring rate was 200 rpm. Following the experiment, the data were integrated, corrected for non-specific heat effects, normalized for the concentration, and analyzed with the assumption of 1:1 binding stoichiometry.

Data Bank accession numbers

Chemical shift assignments have been deposited with BioMagResBank† (accession numbers 5205 for CRT(221–256) and 5204 for CRT(189–261)). The atomic coordinates of the NMR structures of CRT(221–256) (PDB ID code 1K91) and CRT(189–261) (PDB ID code 1K9C) have been deposited with the Protein Data Bank‡.

Acknowledgements

We thank Dr R. Riek for advice on the resonance assignment procedures used for CRT(189–261), Dr K. F. Fischer for help with the RAFT calculations, Dr M. Bouvier for critical reading of the manuscript, Eva Frickel for helpful discussions and preparation of ERp57, and Christiane Schirra for excellent technical assistance. Financial support by the Schweizerischer Nationalfonds (projects 31.51054.97 (A.H.) and 31.49047.96 (K.W.)) and the use of the computing facilities of the Competence Center for Computational Chemistry of the ETH Zürich are gratefully acknowledged.

† www.bmrb.wisc.edu
‡ www.rcsb.org

References

1. Ellgaard, L. & Helenius, A. (2001). ER quality control: towards an understanding at the molecular level. *Curr. Opin. Cell Biol.* **13**, 431–437.
2. Hammond, C., Braakman, I. & Helenius, A. (1994). Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl Acad. Sci. USA*, **91**, 913–917.
3. Ware, F. E., Vassilakos, A., Peterson, P. A., Jackson, M. R., Lehrman, M. A. & Williams, D. B. (1995). The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J. Biol. Chem.* **270**, 4697–4704.
4. Hebert, D. N., Foellmer, B. & Helenius, A. (1995). Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, **81**, 425–433.
5. Parodi, A. J. (2000). Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.* **69**, 69–93.
6. Oliver, J. D., van der Wal, F. J., Bulleid, N. J. & High, S. (1997). Interaction of the thiol-dependent reductase ERp57 with nascent glycoproteins. *Science*, **275**, 86–88.
7. Oliver, J. D., Roderick, H. L., Llewellyn, D. H. & High, S. (1999). ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell*, **10**, 2573–2582.
8. Molinari, M. & Helenius, A. (1999). Glycoproteins form mixed disulphides with oxidoreductases during folding in living cells. *Nature*, **402**, 90–93.
9. Schrag, J. D., Bergeron, J. J., Li, Y., Borisova, S., Hahn, M., Thomas, D. Y. & Cygler, M. (2001). The structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol. Cell*, **8**, 633–644.
10. Ellgaard, L., Riek, R., Herrmann, T., Güntert, P., Braun, D., Helenius, A. & Wüthrich, K. (2001). NMR structure of the calreticulin P-domain. *Proc. Natl Acad. Sci. USA*, **98**, 3133–3138.
11. Frickel, E.-M., Riek, R., Jelesarov, I., Helenius, A., Wüthrich, K. & Ellgaard, L. (2002). TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc. Natl Acad. Sci. USA*, **99**, 1954–1959.
12. Fischer, K. F. & Marqusee, S. (2000). A rapid test for identification of autonomous folding units in proteins. *J. Mol. Biol.* **302**, 701–712.
13. Bouvier, M. & Stafford, W. F. (2000). Probing the three-dimensional structure of human calreticulin. *Biochemistry*, **39**, 14950–14959.
14. Li, Z., Stafford, W. F. & Bouvier, M. (2001). The metal ion binding properties of calreticulin modulate its conformational flexibility and thermal stability. *Biochemistry*, **40**, 11193–11201.
15. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.
16. Ellgaard, L., Riek, R., Braun, D., Herrmann, T., Helenius, A. & Wüthrich, K. (2001). Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment. *FEBS Letters*, **488**, 69–73.
17. Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J. Magn. Reson. ser. B*, **101**, 201–205.
18. Billeter, M., Braun, W. & Wüthrich, K. (1982). Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Computation of sterically allowed proton-proton distances and statistical analysis of proton–proton distances in single crystal protein conformations. *J. Mol. Biol.* **155**, 321–346.
19. Wagner, G. & Wüthrich, K. (1982). Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **155**, 347–366.
20. Herrmann, T., Güntert, P. & Wüthrich, K. (2002). Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol.* **319**, 209–227.
21. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
22. Wüthrich, K. (1994). NMR assignments as a basis for structural characterization of denatured states of globular protein. *Curr. Opin. Struct. Biol.* **4**, 93–99.
23. Perkins, S. J. & Wüthrich, K. (1979). Ring current effects in the conformation dependent NMR chemical shifts of aliphatic protons in the basic pancreatic trypsin inhibitor. *Biochim. Biophys. Acta*, **576**, 409–423.
24. Wüthrich, K. (1976). *NMR in Biological Research: Peptides and Proteins*, North Holland, Amsterdam.
25. Imperiali, B. & Ottesen, J. J. (1999). Uniquely folded mini-protein motifs. *J. Pept. Res.* **54**, 177–184.
26. Macias, M. J., Gervais, V., Civera, C. & Oschkinat, H. (2000). Structural analysis of WW domains and design of a WW prototype. *Nature Struct. Biol.* **7**, 375–379.
27. Kalia, Y. N., Brocklehurst, S. M., Hipps, D. S., Appella, E., Sakaguchi, K. & Perham, R. N. (1993). The high-resolution structure of the peripheral subunit-binding domain of dihydrolipoamide acetyltransferase from the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *J. Mol. Biol.* **230**, 323–341.
28. Spector, S., Young, P. & Raleigh, D. P. (1999). Native-like structure and stability in a truncation mutant of a protein minidomain: the peripheral subunit-binding domain. *Biochemistry*, **38**, 4128–4136.
29. Spector, S., Kuhlman, B., Fairman, R., Wong, E., Boice, J. A. & Raleigh, D. P. (1998). Cooperative folding of a protein minidomain: the peripheral subunit-binding domain of the pyruvate dehydrogenase multienzyme complex. *J. Mol. Biol.* **276**, 479–489.
30. McKnight, C. J., Doering, D. S., Matsudaira, P. T. & Kim, P. S. (1996). A thermostable 35-residue subdomain within villin headpiece. *J. Mol. Biol.* **260**, 126–134.
31. McKnight, C. J., Matsudaira, P. T. & Kim, P. S. (1997). NMR structure of the 35-residue villin headpiece subdomain. *Nature Struct. Biol.* **4**, 180–184.
32. Peng, Z. Y. & Wu, L. C. (2000). Autonomous protein folding units. *Advan. Protein Chem.* **53**, 1–47.
33. Martinez-Yamout, M., Legge, G. B., Zhang, O., Wright, P. E. & Dyson, H. J. (2000). Solution structure of the cysteine-rich domain of the *Escherichia coli* chaperone protein DnaJ. *J. Mol. Biol.* **300**, 805–818.
34. Peterson, J. R. & Helenius, A. (1999). In vitro reconstitution of calreticulin-substrate interactions. *J. Cell Sci.* **112**, 2775–2784.
35. Ellgaard, L., Holtet, T. L., Nielsen, P. R., Etzerodt, M., Gliemann, J. & Thøgersen, H. C. (1997). Dissection of

the domain architecture of the $\alpha_2$macroglobulin-receptor-associated protein. *Eur. J. Biochem.* **244**, 544–551.
36. Bax, A. & Grzesiek, S. (1993). Methodological advances in protein NMR. *Accts Chem. Res.* **26**, 131–138.
37. Bax, A., Clore, G. M. & Gronenborn, A. M. (1990). ^{1}H–^{1}H correlation *via* isotropic mixing of ^{13}C magnetization, a new 3-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. *J. Magn. Reson.* **88**, 425–431.
38. Zerbe, O., Szyperski, T., Ottiger, M. & Wüthrich, K. (1996). Three-dimensional ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]-HMQC for aromatic spin system identification in uniformly ^{13}C-labeled proteins. *J. Biomol. NMR*, **7**, 99–106.
39. Ikura, M., Kay, L. E., Tschudin, R. & Bax, A. (1990). 3-Dimensional NOESY-HMQC spectroscopy of a ^{13}C-labeled protein. *J. Magn. Reson.* **86**, 204–209.
40. Boelens, R., Burgering, M., Fogh, R. H. & Kaptein, R. (1994). Time-saving methods for heteronuclear multidimensional NMR of (^{13}C,^{15}N)-doubly labeled proteins. *J. Biomol. NMR*, **4**, 201–213.
41. Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wüthrich, K. (1983). Improved spectral resolution in COSY ^{1}H-NMR spectra of proteins *via* double quantum filtering. *Biochem. Biophys. Res. Commun.* **117**, 479–485.
42. Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1987). Practical aspects of the E-COSY technique—measurement of scalar spin spin coupling-constants in peptides. *J. Magn. Reson.* **75**, 474–492.
43. Bartels, C., Xia, T. H., Billeter, M., Güntert, P. & Wüthrich, K. (1995). The program XEASY for computer-supported NMR-spectral analysis of biological macromolecules. *J. Biomol. NMR*, **6**, 1–10.
44. Güntert, P., Billeter, M., Ohlenschläger, O., Brown, L. R. & Wüthrich, K. (1998). Conformational analysis of protein and nucleic acid fragments with the new grid search algorithm FOUND. *J. Biomol. NMR*, **12**, 543–548.
45. Nilges, M. (1995). Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.* **245**, 645–660.
46. Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136–146.
47. Koradi, R., Billeter, M. & Güntert, P. (2000). Point-centered domain decomposition for parallel molecular dynamics simulation. *Comput. Phys. Commun.* **124**, 139–147.
48. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M. *et al.* (1995). A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
49. Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
50. Johnson, C. E. & Bovey, F. A. (1958). Calculation of nuclear magnetic resonance spectra of aromatic hydrocarbons. *J. Chem. Phys.* **29**, 1012–1014.
51. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326.

Edited by M. F. Summers

(Received 1 May 2002; received in revised form 26 July 2002; accepted 30 July 2002)