Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment

Lars Ellgaard[a], Roland Riek[b], Daniel Braun[b], Torsten Herrmann[b], Ari Helenius[a,*], Kurt Wüthrich[b,1]

[a] *Institut für Biochemie, Eidgenössische Technische Hochschule, Universitätstrasse 16, CH-8092 Zurich, Switzerland*
[b] *Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule, CH-8093 Zurich, Switzerland*

Received 11 November 2000; revised 27 November 2000; accepted 27 November 2000

First published online 20 December 2000

Edited by Thomas L. James

Abstract Calreticulin (CRT) is an abundant molecular chaperone of the endoplasmic reticulum. Its central, proline-rich P-domain, comprising residues 189–288, contains three copies of each of two repeat sequences (types 1 and 2), which are arranged in a characteristic '111222' pattern. Here we show that the three-dimensional structure of CRT(189–288) contains a single hairpin fold formed by the entire polypeptide chain. The loop at the bottom of the hairpin consists of residues 227–247, and is closed by an anti-parallel β-sheet of residues 224–226 and 248–250. Two additional β-sheets contain residues 207–209 and 262–264, and 190–192 and 276–278. The 17-residue spacing of the β-strands in the N-terminal part of the hairpin and the 14-residue spacing in the C-terminal part reflect the length of the type 1 and type 2 sequence repeats. As a consequence of this topology the peptide segments separating the β-strands in the N-terminal part of the hairpin are likely to form bulges to accommodate the extra residues. These results are based on nearly complete sequence-specific NMR assignments for CRT(189–288), which were obtained using standard NMR techniques with the $^{13}C/^{15}N$-labeled protein, and collection of nuclear Overhauser enhancement upper distance constraints.

Key words: Calreticulin; Nuclear magnetic resonance; Assignment; Topology

1. Introduction

The endoplasmic reticulum (ER) of eukaryotic cells contains a complex machinery for assisted folding of newly synthesized proteins as well as a quality control system to ensure their structural integrity [1,2]. The calcium-binding molecular chaperone calreticulin (CRT) and its close homolog calnexin (CNX) play a pivotal role in these processes. Both proteins are lectins that associate transiently with nascent and newly synthesized glycoproteins in living cells [3–5], by interacting with partially trimmed, monoglucosylated *N*-linked oligosaccharide moieties [6–8]. They co-operate with independently acting enzymes that regulate the cycle of substrate binding to and release from the chaperones [9,10]. Furthermore, they associate with ERp57, a thiol oxidoreductase that is involved in formation of disulfide bonds [11].

CRT (400 amino acids, 46 kDa) is a highly conserved soluble protein (for a review on CRT see [12]). The amino acid sequence shows three regions, the N-, P- and C-domains, with somewhat arbitrarily defined boundaries. The C-domain (residues 286–400) contains the KDEL sequence for ER retrieval. Being highly negatively charged it binds approximately 25 mol Ca^{2+}/mol protein, making CRT one of the major calcium-binding proteins of the ER. Much interest has been focused on the proline-rich domain (P-domain) (residues 171–285), which has a unique, proline-rich sequence with two sequence repeats of 17 (type 1) and 14 residues (type 2), respectively. CRT comprises three copies of each repeat type in a '111222' arrangement (Fig. 4), while CNX has four copies in a '11112222' arrangement. The P-domain has been shown to contain a high affinity Ca^{2+}-binding site [13], and since it is a central part of all truncated constructs that retain binding for glycoproteins [14,15] it has been implicated in substrate binding.

To date, the three-dimensional (3D) structure of neither CRT nor CNX is known, although the crystallization of a CNX ecto-domain fragment has been published [16]. Analytical ultracentrifugation and gel-filtration indicate that CRT is a highly asymmetric molecule ([17]; M. Bouvier, personal communication). Secondary structure prediction programs predict that the P-domain is devoid of regular secondary structure elements, which may at least in part be rationalized by the high content of prolyl residues. Here we describe the 3D structure topology of rat CRT(189–288) determined by nuclear magnetic resonance (NMR) spectroscopy in solution [18].

2. Materials and methods

2.1. Protein expression and purification

The 'pGEX-CRTwt' vector [15] containing the full-length rat CRT cDNA (GenBank accession number X79327) was used as the template in a PCR reaction to generate a fragment encoding the amino acid residues 189–288. The PCR product was cloned into a pRSET A-derived *Escherichia coli* expression vector [19] (kindly provided by Dr. A. Fersht). This vector encodes a N-terminal 17-amino acid tail

*Corresponding author. Fax: (41)-1-632 1269.
E-mail: ari.helenius@bc.biol.ethz.ch

[1] Also corresponding author. Fax: (41)-1-633 1151.

Abbreviations: CNX, calnexin; CRT, calreticulin; ER, endoplasmic reticulum; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser enhancement; NTA, nitrilotriacetic acid; P-domain, proline-rich domain

0014-5793/01/$20.00
PII: S0014-5793(00)02382-6

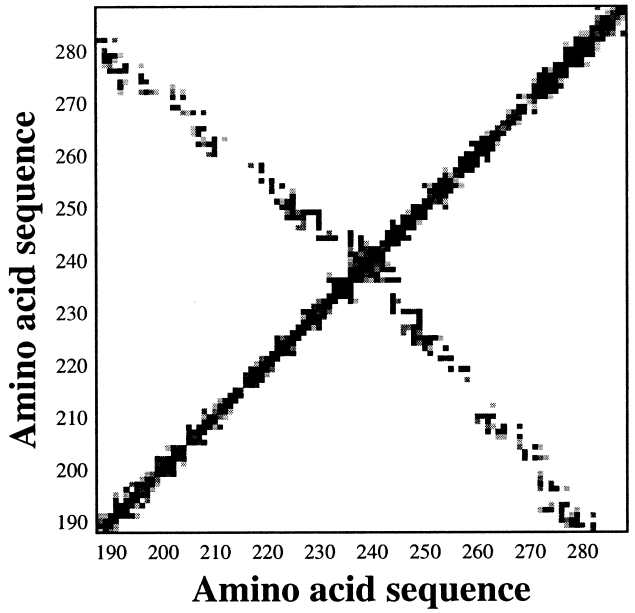

Fig. 1. Diagonal plot of the NOE constraints identified in CRT(189–288). The sequence numbering is shown on both axes. A square indicates the presence of a NOE between a pair of residues. The darkness of the squares increases with increasing number of NOE constraints between the two residues, with black squares representing five or more NOEs. No distinction is made between NOEs involving backbone or side chain hydrogens.

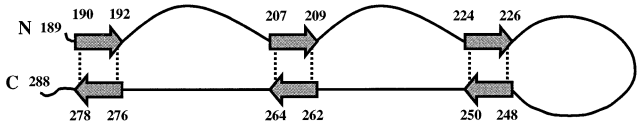

Fig. 2. Topology of the 3D structure of the CRT P-domain. The sequence positions of the first and last residues in the six β-strands are indicated. Vertical dotted lines represent the interstrand $d_{\alpha\alpha}(i,j)$ NOE connectivities across the β-sheets (see Fig. 5).

that includes a hexa-histidine segment and a thrombin cleavage site. After thrombin cleavage, a GS dipeptide segment remains at the N-terminus of the CRT(189–288) construct. The correct sequence of the construct was verified by DNA sequencing.

To produce $^{13}C/^{15}N$-labeled CRT(189–288), a preculture of 65 ml LB medium containing ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) was inoculated with 1 ml overnight culture of *E. coli* BL21(DE3)pLysS cells freshly transformed with expression plasmid. The culture was grown at 37°C until reaching an OD_{600} of 0.8, harvested by centrifugation and resuspended in 50 ml prewarmed M9 minimal medium containing 20 μM $CaCl_2$, 1 mM $MgSO_4$ and antibiotics as described above. The medium was supplemented with 2 g/l ^{13}C-glucose and 1 g/l $^{15}NH_4Cl$ for isotope labeling. The resuspended cells were inoculated into 450 ml M9 medium supplemented as described above and then grown to $OD_{600} = 0.6$. Induction of expression was initiated by adding 1 mM isopropyl L-D-galactopyranoside (IPTG). The cells were harvested after 8 h and resuspended in 30 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 min was applied to a Ni^{2+}-charged nitrilotriacetic acid (NTA) metal–chelate affinity column (15 ml), which was washed thoroughly 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 digesting with 2.5 units of thrombin per mg of fusion protein for 24 h at room temperature. The digest 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 400 mM NaCl. Finally, the protein was gel-filtrated into a 10-mM $CaCl_2$ solution and concentrated using Centricon spin columns (Amicon). Pure, uniformly $^{13}C/^{15}N$-labeled CRT(189–288) was obtained in yields of 7 mg per liter of M9 minimal medium. For preparation of the NMR sample in D_2O solution, the H_2O solution was replaced by successive concentration and dilution steps. The correct molecular mass of the protein was verified by matrix-assisted laser desorption and ionization time of flight (MALDI-TOF) mass spectrometry. Since no activity assay exists, the structural integrity of the refolded CRT(189–288) was probed by circular dichroism (CD) spectroscopy (data not shown), which revealed a co-operative thermal denaturation transition of the protein with a mid-point temperature of approximately 35°C.

2.2. NMR spectroscopy and structure determination

All NMR spectra were acquired with uniformly $^{13}C/^{15}N$-labeled CRT(189–288) at a protein concentration of 3 mM in 95% H_2O/5% D_2O and at 1.5 mM protein concentration in 100% D_2O, pH = 6.3, $T = 20$°C, 10 mM $CaCl_2$. The NMR measurements were performed on Bruker DRX600, DRX750 and DRX800 spectrometers equipped with four radio-frequency channels and triple resonance probeheads with shielded *z*-gradient coils. For the resonance assignments and collection of conformational constraints the following experiments were recorded: 3D HNCA [20], 3D CBCA(CO)NH [20], 3D *ct*-HCCH-TOCSY with 14 ms mixing time [21], 3D ^{1}H-TOCSY-relayed ct-[^{13}C,^{1}H]HMQC [22], 3D combined $^{15}N/^{13}C$-resolved [^{1}H,^{1}H]-nuclear Overhauser enhancement spectrometry (NOESY) [23,24] in H_2O, $^{13}C^R$-resolved [^{1}H,^{1}H]-NOESY [23] in H_2O, and ^{13}C-resolved [^{1}H,^{1}H]-NOESY [23] in D_2O. NOESY spectra were recorded with mixing times $\tau_m = 60$ ms. The ^{1}H, ^{15}N and ^{13}C chemical shifts are

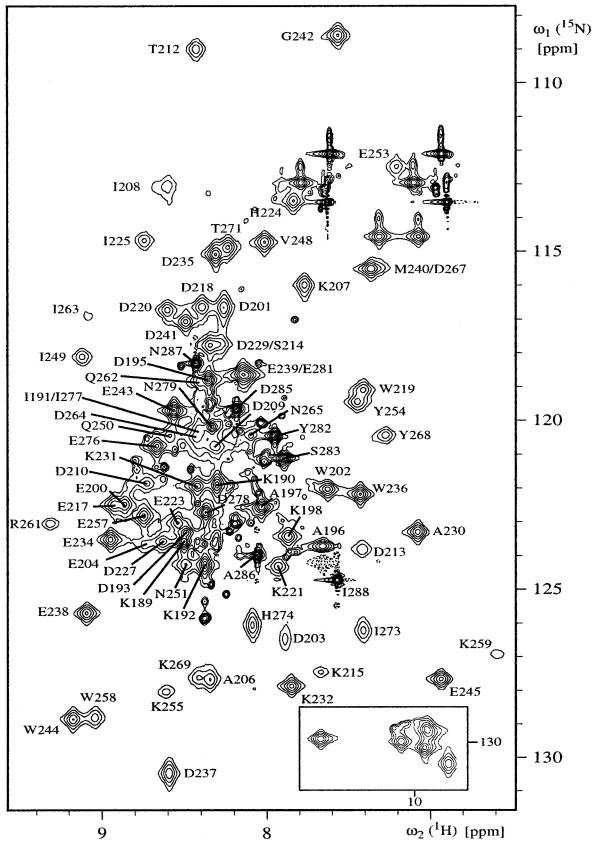

Fig. 3. 2D [^{15}N, ^1H]-COSY spectrum of CRT(189–288) with backbone ^{15}N–^1H resonance assignments. The insert in the lower right contains the indole $^{15}N^{\varepsilon}$–^1H cross peaks of the six tryptophanyl residues. The sequence numbering is that of rat CRT, GenBank accession number CAA55890.

calibrated relative to 2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt (DSS). The spectra were processed with the program PROSA [25] and analyzed with the program XEASY [26]. Initial structure calculations were performed using a new version of the program DYANA [27] integrated with a new algorithm for automated NOE cross peak assignment, CANDID (T. Herrmann, P. Güntert and K. Wüthrich, unpublished).

3. Results and discussion

The diagonal plot of the NOE constraints in Fig. 1 contains a 'double diagonal' pattern, where the diagonal going from bottom left to top right of the plot shows sequential and medium-range upper distance constraints, and the second, approximately perpendicular diagonal shows long-range constraints [18]. This pattern of NOE constraints shows that the P-domain forms an extended hairpin fold, where the two sides of the hairpin run anti-parallel to each other, the N- and C-termini are in close proximity, and the hairpin does not fold back on itself (Fig. 2). In the following we describe the NMR experiments that resulted in the characterization of the global fold of CRT(189–288) and some further details of the molecular architecture, as surveyed in Fig. 4.

3.1. NMR assignment of CRT(189–288)

The NMR data in Fig. 1 and the schematic representation of the polypeptide fold in Fig. 2 are based on nearly complete sequence-specific resonance assignments of CRT(189–288). These are surveyed in the 2D [^{15}N, ^{1}H]-COSY spectrum of CRT(189–288) in Fig. 3, and in a plot of the key data versus the amino acid sequence in Fig. 4. The sequential assignment was based on standard triple resonance spectra and on sequential NOEs [20,28,29]. Nearly complete sequential assignments were eventually obtained from a HNCA spectrum [20] in combination with sequential NOE connectivities [28,29] (Fig. 4). Out of 84 backbone amide resonances expected in the 2D [^{15}N, ^{1}H]-COSY spectrum we identified 81 (Fig. 3). The three missing resonances are those of Arg205, Gly256 and Gly270, but we were able to observe Gly256 in the 3D combined ^{15}N/^{13}C-resolved [^{1}H, ^{1}H]-NOESY spectrum recorded in H_2O. Out of 16 Xxx–Pro dipeptide segments, 14 were identified based on sequential $d_{\alpha\delta}$ and $d_{\beta\delta}$ NOE connectivities [18]. The chemical shift lists have been deposited in the BioMagResBank (accession number 4878).

In the 2D [^{15}N, ^{1}H]-COSY spectrum (Fig. 3) some additional, so far unassigned weak peaks are observed, mostly located next to strong signals from flexible residues near the C-terminus. These may reflect the presence of minor populations of one or several conformational isoforms of CRT(189–288), which might typically arise from *cis*/*trans* isomerization of Xxx–Pro peptide bonds [18].

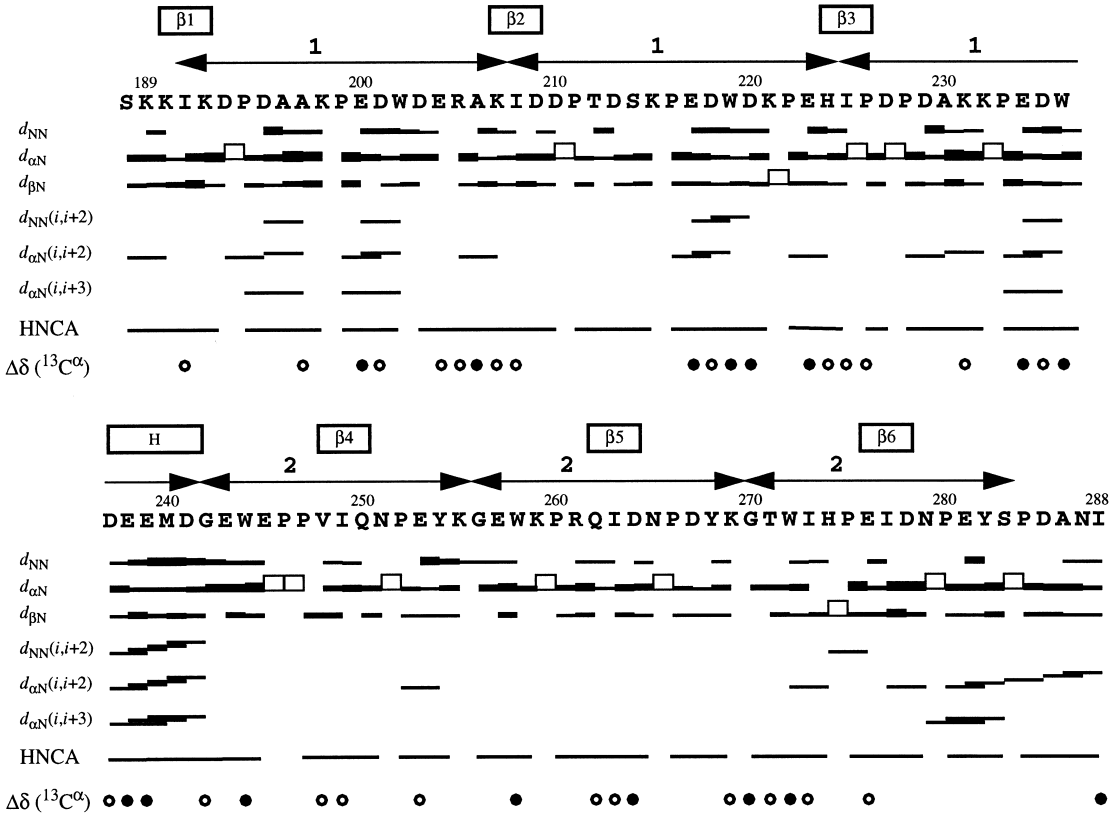

Fig. 4. Survey of sequential and medium-range NOE connectivities, sequential connectivities observed in a HNCA spectrum, and deviation of observed $^{13}C^{\alpha}$ chemical shifts from the random coil values for CRT(189–288). The positions of the six β-strands (β1–β6) and a helix-like region (H) as deduced from the NOESY data are indicated above the sequence. The arrows above the sequence indicate the positions of the type 1 and type 2 repeats in the CRT sequence. For the sequential NOE connectivities d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$ the thickness of the line is proportional to the NOE intensity. Open bars in the $d_{\alpha N}$ and $d_{\beta N}$ rows indicate sequential $d_{\alpha\delta}$ and $d_{\beta\delta}$ NOEs for Xxx–Pro dipeptide segments. Medium-range NOEs are represented by lines connecting the two interacting residues. In the Δδ ($^{13}C^{\alpha}$) row the difference between corresponding observed and random coil chemical shifts is indicated by filled and open circles for residues with Δδ ($^{13}C^{\alpha}$) > 1.0 ppm and Δδ ($^{13}C^{\alpha}$) < −1.0 ppm, respectively. All residues preceding a prolyl residue showed Δδ ($^{13}C^{\alpha}$) < −1.0 ppm, which is not indicated in the figure [34]. The N-terminal seryl residue is not part of the native CRT sequence.

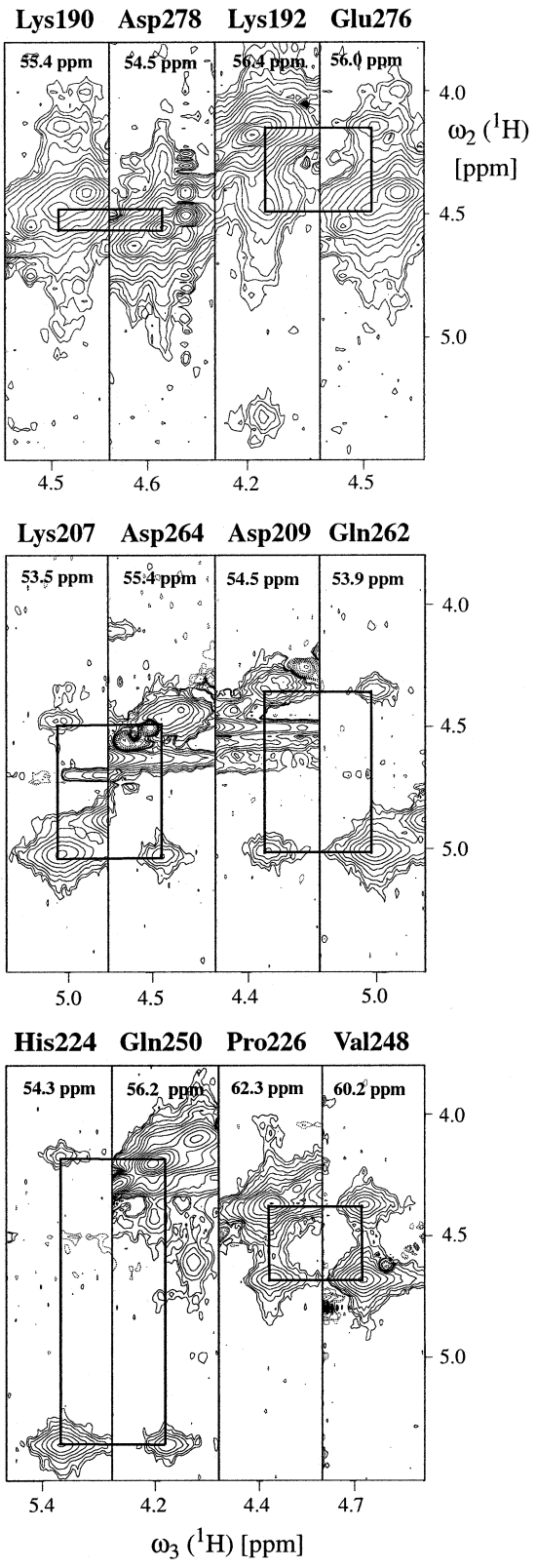

Fig. 5. Identification of three anti-parallel β-sheets in CRT(189–288). Spectral regions from 3D ^{13}C-resolved [^{1}H, ^{1}H]-NOESY spectra recorded either in H_2O solution (residues Lys190, Lys192, Glu276, Asp278) or in D_2O solution (residues Lys207, Asp209, His224, Pro226, Val248, Gln250, Gln262, Asp264) are shown. Squares connect the cross peaks and the diagonal positions for the $d_{\alpha\alpha}(i,j)$ connectivities between the residue pairs shown.

3.2. Regular secondary structures and topology of CRT(189–288)

The data in Figs. 1 and 4 provide only little information on regular secondary structures in CRT(189–288), which may to a large part be due to the high Pro content. A short helix-like structure near E239 was identified based on medium-range NOEs (see Fig. 4). As judged from the presence of $d_{\alpha N}(i,i+2)$ connectivities, this segment shows characteristics of a 3_{10} helix [30]. In principle, the global hairpin arrangement of the polypeptide chain could accommodate extensive formation of anti-parallel β-sheet structure. However, long-range H^{α}-H^{α} NOEs indicative of anti-parallel β-sheets [18] were found in the 3D ^{13}C-resolved [^{1}H, ^{1}H]-NOESY spectra only between the residue pairs (190, 278), (192, 276), (207, 264), (209, 262), (224, 250) and (226, 248) (Fig. 5), which define the short β-sheets depicted in Fig. 2. The presence of these β-sheets was supported by further NOE constraints, and is compatible with the $^{13}C^{\alpha}$ chemical shift data (Fig. 4) [31,32].

As shown in Fig. 4, the position of the β-strands β1, β2 and β3 is the same within each type 1 repeat sequence, and the position of the β-strands β4, β5 and β6 is the same within each type 2 repeat sequence. In the 3D structure topology of CRT(189–288) (Fig. 2), this leads to equidistant spacing of the β-strands within each side of the hairpin, but different spacings on the two sides. Thus, the distances between β1 and β2, and between β2 and β3 is 17 residues, which reflects the length of the type 1 repeat sequence (see Fig. 4). Likewise, the length of the type 2 repeat sequence is reflected in the 14-residue distances between β4 and β5, and between β5 and β6. Therefore, the repetitive nature of the P-domain sequence also seems to be maintained on the 3D structure level, which suggests the formation of 'bulges' to accommodate the extra three residues of the peptide segments between β1 and β2, and between β2 and β3 (Fig. 2). By extension of this topology, the CNX P-domain, which has an additional set of sequence repeats, would also contain an additional β-sheet when compared with CRT.

With the N- and C-termini located in close proximity to each other, the P-domain could constitute a finger-like extension from the CRT core structure. This could explain the unusual hydrodynamic properties of the protein. Whereas it is possible that this structure could provide a binding site for glycans, it seems more likely to be a site for protein–protein interactions, for instance involving the co-chaperone ERp57 with which both CNX and CRT form a complex [11]. Alternatively, the P-domain could be the recently described site of direct interaction with misfolded proteins [33].

In conclusion, the CRT(189–288) fragment shows a unique global fold, which seems to be intimately related with long-recognized features of the primary structure. The ongoing refinement of the solution NMR structure (L. Ellgaard, R. Riek, T. Herrmann, P. Güntert, D. Braun, A. Helenius and K. Wüthrich, to be published) should reveal the intramolecular contacts that stabilize this polypeptide fold, in addition to the three short β-sheets shown in Fig. 2.

Acknowledgements: Financial support by the Schweizerischer Nationalfonds (Projects 31.51054.97 (A.H.), 31.49047.96 and 438+050287 (K.W.)) is gratefully acknowledged.

References

[1] Gething, M.J. and Sambrook, J. (1992) Nature 355, 33–45.
[2] Ellgaard, L., Molinari, M. and Helenius, A. (1999) Science 286, 1882–1888.
[3] Hochstenbach, F. (1992) Hum. Cell 5, 12–24.
[4] Ou, W.J., Cameron, P.H., Thomas, D.Y. and Bergeron, J.J. (1993) Nature 364, 771–776.
[5] Peterson, J.R., Ora, A., Van, P.N. and Helenius, A. (1995) Mol. Biol. Cell 6, 1173–1184.
[6] Hammond, C., Braakman, I. and Helenius, A. (1994) Proc. Natl. Acad. Sci. USA. 91, 913–917.
[7] Spiro, R.G., Zhu, Q., Bhoyroo, V. and Soling, H.D. (1996) J. Biol. Chem. 271, 11588–11594.
[8] Ware, F.E., Vassilakos, A., Peterson, P.A., Jackson, M.R., Lehrman, M.A. and Williams, D.B. (1995) J. Biol. Chem. 270, 4697–4704.
[9] Helenius, A., Trombetta, E.S., Hebert, D.N. and Simons, J.F. (1997) Trends Cell Biol. 7, 193–200.
[10] Bergeron, J.J., Zapun, A., Ou, W.J., Hemming, R., Parlati, F., Cameron, P.H. and Thomas, D.Y. (1998) Adv. Exp. Med. Biol. 435, 105–116.
[11] Oliver, J.D., Roderick, H.L., Llewellyn, D.H. and High, S. (1999) Mol. Biol. Cell 10, 2573–2582.
[12] Michalak, M., Corbett, E.F., Mesaeli, N., Nakamura, K. and Opas, M. (1999) Biochem. J. 344, 281–292.
[13] Baksh, S. and Michalak, M. (1991) J. Biol. Chem. 266, 21458–21465.
[14] Vassilakos, A., Michalak, M., Lehrman, M.A. and Williams, D.B. (1998) Biochemistry 37, 3480–3490.
[15] Peterson, J.R. and Helenius, A. (1999) J. Cell Sci. 112, 2775–2784.
[16] Hahn, M., Borisova, S., Schrag, J.D., Tessier, D.C., Zapun, A., Tom, R., Kamen, A.A., Bergeron, J.J.M., Thomas, D.Y. and Cygler, M. (1998) J. Struct. Biol. 123, 260–264.
[17] Waisman, D.M., Salimath, B.P. and Anderson, M.J. (1985) J. Biol. Chem. 260, 1652–1660.
[18] Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York.
[19] Zahn, R., Buckle, A.M., Perrett, S., Johnson, C.M., Corrales, F.J., Golbik, R. and Fersht, A.R. (1996) Proc. Natl. Acad. Sci. USA 93, 15024–15029.
[20] Bax, A. and Grzesiek, S. (1993) Acc. Chem. Res. 26, 131–138.
[21] Bax, A., Clore, G.T.M. and Gronenborn, A.M. (1990) J. Magn. Reson. 88, 425–431.
[22] Zerbe, O., Szyperski, T., Ottiger, M. and Wüthrich, K. (1996) J. Biomol. NMR 7, 99–106.
[23] Ikura, M., Kay, L.E., Tschudin, R. and Bax, A. (1990) J. Magn. Reson. 86, 204–209.
[24] Boelens, R., Burgering, M., Fogh, R.H. and Kaptein, R. (1994) J. Biomol. NMR 4, 201–213.
[25] Güntert, P., Dötsch, V., Wider, G. and Wüthrich, K. (1992) J. Biomol. NMR 2, 619–629.
[26] Bartels, C., Xia, T., Billeter, M., Güntert, P. and Wüthrich, K. (1995) J. Biomol. NMR 6, 1–10.
[27] Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) J. Mol. Biol. 273, 283–298.
[28] Billeter, M., Braun, W. and Wüthrich, K. (1982) J. Mol. Biol. 155, 321–346.
[29] Wagner, G. and Wüthrich, K. (1982) J. Mol. Biol. 155, 347–366.
[30] Wüthrich, K., Billeter, M. and Braun, W. (1984) J. Mol. Biol. 180, 715–740.
[31] Spera, S. and Bax, A. (1991) J. Am. Chem. Soc. 113, 5490–5492.
[32] Luginbühl, P., Szyperski, T. and Wüthrich, K. (1995) J. Magn. Reson. B 109, 229–233.
[33] Saito, Y., Ihara, Y., Leach, M.R., Cohen-Doyle, M.F. and Williams, D.B. (1999) EMBO J. 18, 6718–6729.
[34] Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) J. Biol. NMR 5, 67–81.