Chapter 2 Structural Analysis of Calreticulin, an Endoplasmic Reticulum-Resident Molecular Chaperone



Gunnar Houen, Peter Højrup, Evaldas Ciplys, Christine Gaboriaud, and Rimantas Slibinskas

Abstract Calreticulin (Calr) is an endoplasmic reticulum (ER) chaperone involved in protein quality control, Ca²⁺ regulation and other cellular processes. The structure of Calr is unusual, reflecting different functions of the protein: a proline-rich β -hairpin arm and an acidic C-terminal tail protrude from a globular core, composed of a β -sheet sandwich and an α -helix. The arm and tail interact in the presence of Ca²⁺ and cover the upper β -sheet, where a carbohydrate-binding site gives the chaperone glycoprotein affinity. At the edge of the carbohydrate-binding site is a conserved, strained disulphide bridge, formed between C¹⁰⁶ and C¹³⁷ of human Calr, which lies in a polypeptide-binding site. The lower β -sheet has several conserved residues, comprised of a characteristic triad, D¹⁶⁶-H¹⁷⁰-D¹⁸⁷, Tyr¹⁷² and the free C¹⁶³. In addition to its role in the ER, Calr translocates to the cell surface upon stress and functions as an immune surveillance marker. In some myeloproliferative neoplasms, the acidic Ca²⁺-binding C-terminal tail is transformed into a polybasic sequence.

G. Houen (🖂)

Institute of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark

e-mail: gunnar.houen@regionh.dk

P. Højrup Institute of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark e-mail: php@bmb.sdu.dk

E. Ciplys · R. Slibinskas Department of Eucaryote Gene Engineering, Life Sciences Center, Vilnius University, Vilnius, Lithuania e-mail: evaldas.ciplys@bti.vu.lt; rimantas.slibinskas@bti.vu.lt

C. Gaboriaud Institut de Biologie Structurale Jean-Pierre Ebel, CEA, Grenoble, France e-mail: christine.gaboriaud@ibs.fr

© Springer Nature Switzerland AG 2021 L. B. Agellon, M. Michalak (eds.), *Cellular Biology of the Endoplasmic Reticulum*, Progress in Molecular and Subcellular Biology 59, https://doi.org/10.1007/978-3-030-67696-4_2

Department of Neurology, Glostrup Research Institute, Glostrup, Denmark

Keywords Calreticulin \cdot Calnexin \cdot Chaperone \cdot Lectin \cdot Protein synthesis \cdot Protein stability \cdot Protein structure

Abbreviations

Calr	Calreticulin
Calr3	Calsperin
Canx	Calnexin
CDS	Circular dichroism spectroscopy
Clgn	Calmegin
DTT	Dithiothreitol
ECM	Electron cryo-microscopy
ER	Endoplasmic reticulum
INDELS	Insertions and deletions
JAK	Janus kinase
MHC	Major histocompatibility complex
MM	Molecular modelling
MPN	Myeloproliferative neoplasms
NMRS	Nuclear magnetic resonance spectroscopy
SAXS	Small angle X-ray scattering
XL-MS	Cross-linking/mass spectrometry
XRC	X-ray crystallography

2.1 Introduction

In eukaryotic cells, ER is the site for synthesis, maturation and quality control of secreted, membrane-bound and ER-resident proteins (Halperin et al. 2014; Kepp and Galluzzi 2020). This process involves translation of mRNAs on ER membranebound ribosomes, translocation of the nascent polypeptide across the ER membrane and co-translational folding to the native three-dimensional structure. To assist in these processes, the ER contains elaborate systems involved in post-translational modification and quality control of newly synthesized proteins (Vincenz-Donnelly L and Hipp MS 2017; Halperin et al. 2014). An important protein quality control system in the ER is constituted by the chaperones calreticulin (Calr) and calnexin (Canx), which act in concert with protein disulfide isomerases (e.g. Pdia3/ERp57) to facilitate the synthesis and maturation of proteins, and to prevent accumulation of misfolded proteins (Bergeron et al. 1994; Michalak et al. 1998; Coe and Michalak 2010; Halperin et al. 2014; Matsusaki et al. 2020). Canx is a transmembrane protein, with its major part in the ER lumen, while Calr is an ER luminal protein (Bergeron et al. 1994; Wada et al. 1995; Michalak et al. 1998; Danilczyk et al. 2000). In humans, these proteins are encoded by separate genes with 9 (Calr) or 20 (Canx) exons (McCauliffe et al. 1992; Tjoelker et al. 1994). In mammals, Calr and Canx furthermore play important roles in the loading of peptides on MHC class I (Blees et al. 2017). This chapter reviews current knowledge on the structure of human Calr (hCalr).

2.2 Primary Structures of Calr and Related Proteins

Analyses of Canx and Calr amino acid sequences have revealed that they are evolutionarily conserved proteins encoded by single genes. In the testicles a Canx homologue, calmegin (Clgn) and a Calr homologue, calsperin (Calr3) are also present (Ohsako et al. 1994; Watanabe et al. 1994; Ikawa et al. 2011). In plants, several Calr isoforms are encoded by separate genes (Jia et al. 2009; Del Bem 2011; Wasag et al. 2019). In yeasts, a Canx homologue is present (de Virgilio et al. 1993; Parlati et al. 1995).

Figure 2.1 shows an alignment of amino acid sequences of selected Calr/Canx family members. Common to both Canx and Calr is an N-terminal domain of approximately 200 amino acid residues, followed by a proline-rich domain of approximately 100 (Calr and Calr3) or 145 (Canx and Clgn) amino acid residues, which is followed by a C-terminal domain. In Calr (and Calr3) the C-terminal (C) domain is highly acidic and consists of approximately 100 amino acids ending in an ER retrieval sequence of four residues (KDEL in hCalr). In Canx (and Clgn) the C-terminal domain has a transmembrane stretch of 20 amino acids and a cytoplasmic "tail" of approximately 90 amino acids.

Relatively few residues are invariant among all Calr and related proteins (Fig. 2.1) and these mainly reside in the N-terminal (N) domain and the middle proline-rich (P) domain. Conserved stretches/clusters of amino acids in the N-domain are found in the vicinity of the conserved disulfide bridge ($C^{106}-C^{137}$ in Calr) and in the P-domain in the neighborhood of conserved W residues, being part of characteristic repeat sequences. A third interesting residue (C^{163} in hCalr) is conserved in all known mammalian and some other Calrs.

2.3 Higher Order Structures of Calr and Related Proteins

Table 2.1 shows a list of Calr and Canx proteins from different species with solved three dimensional (3D) structures. Figure 2.2 shows two partial structures for human Calr, plus one view of its full-length elongated structure in the context of the MHC-I peptide-loading complex. The 3D structures of Canx and Calr show close similarity but also some important differences. Both proteins have a globular core consisting of a β -sheet sandwich, where the strands are arranged in a "jelly roll" fold, and an α -helix. The sandwich is composed of two β -sheets, where the N-terminal 200 amino acids contribute 6 strands to the first, "upper" β -sheet and 6 strands to the second "lower" β -sheet. Intertwined in the two sheets are two β -strands derived from the

А						M	FTLFLLIALS	SAKVYFHETF
В					M	RLLLCLIFLV	FVFNFALSTV	HFKDTFDNDW
C					M	LI SVPLLI GL	I GLAAADPAT	YEKEOFI DGD
D					M	LESVPLLEG	Ι ΟΙ ΔΔΔΠΡΔΤ	YEKEOELDCD
F					M			VEKEOFI DCD
F						===MKSLCLL		VEKEEENDAS
ċ					M			VENEELDCE
U U						MDAATEECA		
Ŧ	MECKWILLOMI							VEADSEDDCT
1	MEGRWELCME	LVLGTTIVQA				TADDCCDV//T	VKADVDTCEV	
J	MUEDAEW	LULGIAIVEA	NACEMODDVID	TEDEEENSEE		IAFF33FKVI	VKTPOPTCEV	
ĸ	MINFQAFW	LCLGLLFISI	NAEFMUDUVE	IEDFEENSEE	TDAM	-ESELSSEIK	TKIPQPIGEV	TFAETFDSGR
								<u>^</u>
			TV 6.61/					
A	ENRUKWIDST	SSGRALGPER	LVSGK	WYGD-ANNKG	LUISEDNKFY	IAAAKLDEEF	SNKDKNLIVQ	YNLKFEQGID
R	ESRWVVSDWH	KEDGKSGKLV	HIAGK	WFGD-ENQKG	LQTSEDARFY	AVSAKEP-SE	SNKGKDLVLQ	YTVKNEQKVD
C	AWINRWVESK	HKSD-FGKFV	LSSGK	FYGDLEKDKG	LQTSQDARFY	ALSAKFEP-F	SNKGQTLVVQ	FIVKHEQNID
D	AWINRWVESK	HKSD-FGKFV	LSSGK	FYGDQEK DKG	LQISQDARFY	ALSARFEP-F	SNKGQILVVQ	FIVKHEQNID
E	GWTSRWIESK	HKSD-FGKFV	LSSGK	FYGDEEK DKG	LQTSQDARFY	ALSASFEP-F	SNKGQTLVVQ	FTVKHEQNID
F	-WEKRWVQSK	HK D D-FGAFK	LSAGK	FFDVESRDQG	IQTSQDAKFY	SRAAKFDKDF	SNKGKTLVIQ	YTVKHEQGID
G	HWRNRWLQST	NDSR-FGHFR	LSSGK	FYGHKEK DKG	LQTTQNGRFY	AISARFKP-F	SNKGKTLVIQ	YTVKHEQKMD
н	KSMEHWTTSK	HR D D-FGKVE	ISAGK	FYADAEK skg	LRLTEDARFY	ALSTAFPTPI	TNEKKSLVVS	FSVKHEQDLK
Ι	LSGWILSKAK	KD D TDDEIAK	YDGKWEVDEM	KETKLPG DKG	LVLMSRAKHH	AISAKLNKPF	LFDT K P LIVQ	YEVNFQNGIE
J	LSGWILSKAK	KD D TDDEIAK	YDGKWEVEEM	KESKLPG DKG	LVLMSRAKHH	AISAKLNKPF	LFDTKP LIVQ	YEVNFQNGIE
Κ	LAGWVLSKAK	KD D MDEEISI	YDGRWEIEEL	KENQVPG DRG	LVLKSRAKHH	AISAVLAKPF	IFADKPLIVQ	YEVNFQDGID
				. : *	* : : .		****	. : :: :: :
A	CGGGYIKLLP	KKSIESEEK F	TPESEYNIMF	GPDVCGG-SK	RT HVI M N YKG	KNNLIRKEI-	КС	ESDDISHLYT
A B	CGGGYIKLLP CGGSYIKLLP	KKSIESEEK F SKLDQSA F	TPESEYNIMF DGESEYSIMF	GPDVCGG-SK GPDVCGA-SK	RT HVI MNYKG RV HVI LNYKG	KNNLIRKEI- KNHLIKKEIN	KC KV	ESDDISHLYT ETDQLTHQYT
A B C	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP	KKSIESEEK F SKLDQSA F SGLDQKD M	TPESEYNIMF DGESEYSIMF HGDSEYNIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK	RT HVI MNYKG RV HVI LNYKG KV HVI FNYKG	KNNLIRKEI- KNHLIKKEIN KNVLINKDI-	KC KV RC	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT
A B C D	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI-	KC KV RC RC	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT
A B C D E	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG KVHVIFNYKG	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- KNVLINKDI-	KC RV RC RC RC	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KDDEFTHLYT
A B C D E F	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKVMR	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- KNVLINKDI- ENKLIKKEI-	KC KV RC RC RC	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT
A B C D E F G	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKVMR CGGGYIKVFP	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF NGKSQYYIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGFDIK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVIFNYKG RVHVILNYKG KVHVILHFKN	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI-	KC RC RC RC TC RC	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT
A B C D E F G H	CGGGYIKLLP CGGSYIKLEP CGGGYVKLFP CGGGYVKLFP CGGGYVKVMR CGGGYIKVFP CGGGYIKLP	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF NGKSQYYIMF HGETKYWLMF	GPDVCGG-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGFDIK GPDICGFDIK GPDRCGS-QN	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVIFNYKG KVHVILNYKG RVHVILNFKN RVHILHYNG	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI-	KC RC RC RC RC RC RF	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT PEDKLTHVYT
A B C D E F G H I	CGGGYIKLLP CGGSYIKLP CGGGYVKLFP CGGGYVKLFP CGGGYVKVFP CGGGYIKVFP CGGGYIKLLP CGGAYVKLLS	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF NGKSQYYIMF HGETKYWLMF HDKTPYTIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGF-DIK GPDRCGS-QN GPDKCGEDYK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILHFKN RVHILHFKN RVHILHFKN	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGVYEEKHA	KC RC RC RC RC RF KRPDADLKTY	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT PEDKLTHVYT FTDKKTHLYT
A B C D E F G H I J	CGGGYIKLLP CGGSYIKLFP CGGGYVKLFP CGGGYVKLFP CGGGYKKVFP CGGGYIKVFP CGGGYIKVFP CGGGYIKLLP CGGAYVKLLS CGGAYVKLLS	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETKYWLMF HDKTPYTIMF HDKTPYTIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGFDIK GPDRCGS-QN GPDKCGEDYK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILHFKN RVHILHYNG LHFIFRHKNP LHFIFRHKNP	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENREWSKRI- KTGVYEEKHA	KC RC RC RC RF KRPDADLKTY KRPDADLKTY	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KDDEFTHLYT KVDGFTHLYT PEDKLTHVYT FTDKKTHLYT
ABCDEFGHIJK	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKVMR CGGGYIKVFP CGGGYIKLLP CGGAYVKLLS CGGAYIKLLA	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDDLILENF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETEYNVMF NGKSQYYIMF HGETKYWLMF HDKTPYTIMF YDKTSYIIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGF-TR GPDRCGS-QN GPDKCGEDYK GPDKCGEDYK	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG RVHVIFNYKG RVHVILNYKG RVHVILHFKN RVHIILHYNG LHFIFRHKNP LHFIFRHKNP LHFIFRHKHP	KNNLIRKEI- KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGVYEEKHA KTGVYEEKHA	KC RC RC RC RC RF KRPDADLKTY KRPDADLKTY KPPDVDLKKF	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT PEDKLTHVYT FTDKKTHLYT FTDRKTHLYT
A B C D E F G H I J K	CGGGYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKVFP CGGGYIKLLP CGGGYIKLLS CGGAYVKLLS CGGAYVKLLS	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF KTPELNLDQF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETEYNVMF HGETEYNVMF HDKTPYTIMF HDKTPYTIMF YDKTSYIIMF YDKTSYIIMF	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TK GPDCGEPIK GPDRCGE-QN GPDKCGEDYK GPDKCGEDYK *** *	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNFKN RVHILHFKN LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGYYEEKHA KTGYYEEKHA KTGYFEEKHA	KC RC RC RC RC RC KRPDADLKTY KRPDADLKTY KPPDVDLKKF	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT FTDKKTHLYT FTDKKTHLYT FTDKKTHLYT
A B C D E F G H I J K	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKVMR CGGGYIKVP CGGGYIKLP CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDLILENF	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF NGKSQYYIMF HGETRYNVMF HDKTPYTIMF HDKTPYTIMF YDKTSYIIMF :: * :**	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPOTK GPDICGPOIK GPDRCGS-QN GPDKCGEDYK GPDKCGEDYK *** **	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILHYKN RVHILHYNG LHFIFRHKNP LHFIFRHKNP LHFIFRHKP	KNNLIRKEI- KNVLINKDI- KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGYYEEKHA KTGYYEEKHA	KC RC RC RC RC RC RC KRPDADLKTY KRPDADLKTY KPPDVDLKKF	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT PEDKLTHVYT FTDKKTHLYT FTDKKTHLYT * * * *
A B C D E F G H I J K A	CGGGYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLP CGGGYIKLP CGGGYVKLLS CGGAYVKLLS CGGAYIKLLA ***.*:: LIIRPNNTYV	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF DTDDLILENF : VKIDGVEKQE	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETPYNVMF HDKTPYTIMF HDKTPYTIMF YDKTSYIIMF YDKTSYIIMF GKFDEDWD	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGFDIK GPDKCGEDYK GPDKCGEDYK CPDKCGEDYK *** ** MLAPKEIDDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG KVHVILHYKG KVHVILHFKN RVHILHYKG LHFIFRHKNP LHFIFRHKNP LHFIFRHKP	KNNLIRKEI- KNNLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGYYEEKHA KTGYFEEKHA EKEIDDPNDK	KC RC RC RC RC RC KRPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KODEFTHLYT KVDGFTHLYT KVDGFTHLYT PEDKLTHLYT FTDKKTHLYT FTDRKTHLYT * * * *
A B C D E F G H I J K A B	CGGGYIKLLP CGGSYIKLP CGGGYVKLFP CGGGYVKLFP CGGGYVKVFP CGGGYIKVFP CGGGYIKVFP CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS X**.*:: LIIRPNNTYV LVISPDNTYK	KKSIESEEKF SKLDQSAF SGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF TDDLILENF DTDDLILENF UKIDGVEKQE VLVDNKEIQA	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETEYNNF HGETEYNNF HGETEYWLMF HDKTPYTIMF HDKTPYTIMF YDKTSYIIMF YDKTSYIIMF GKFDEDWD GNLADDWE	GPDVCGG-SK GPDUCGA-SK GPDICGPCTK GPDICGPCTK GPDICGPCTR GPDICGPCTR GPDICGCPTR GPDRCGC-QN GPDKCGEDYK GPDKCGEDYK K*** ** MLAPKEIDDP LLPSKQIKDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILHYNG LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP 	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGVYEEKHA KTGIYEEKHA : EKEIDDPNDK VKEIDDPEDV	KC RC RC RC RF KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEGWDDIP- KPAGHDDIP-	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDCFTHLYT FTDKKTHLYT FTDKKTHLYT FTDKKTHLYT * * **
A B C D E F G H I J K A B C	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKUFP CGGGYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS LIRPNNTYV LVISPDNTYK LIVRPDNTYK	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDDLILENF VKIDQVEKQE VKIDNSQVES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETRYNVMF HDETPYTIMF HDKTPYTIMF HDKTPYTIMF CKFDEDWD GNLADDWE GSLEDDWD	GPDVCGG-SK GPDICGPGTK GPDICGPGTK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDRCGS-QN GPDKCGEDYK GPDKCGEDYK *** ** MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILHFNYKG HFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKP NVSKPADWVD DAAKPE0WDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENREWSKRI- ENREWSKRI- KTGYYEEKHA KTGYFEKHA EKEIDDPNDK VKEIDDPPDV RAKIDDPTDS	KC RC RC RC RC RC KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEGWDDIP- KPEGWDDIP- KPEDWDKP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KSDELTHLYT PEDKLTHVYT FTDKKTHLYT FTDKKTHLYT * :* **
A B C D E F G H I J K A B C D	CGGGYIKLLP CGGSYIKLP CGGGYVKLFP CGGGYVKLFP CGGGYVKMRR CGGGYIKLP CGGGYIKLLS CGGAYVKLS CGGAYVKLS CGGAYVKLS LIIRPNNTYV LVISPDNTYK LIVRPDNTYE LIVRPDNTYE	KKSIESEEKF SKLDQSAF SGLDQKDM GGLDQKDM NSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDDLILENF VKIDGVEKQE VLIDNKEIQA VKIDNSQVES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF NGKSQYYIMF HGETRYNWMF HDKTPYTIMF HDKTPYTIMF YDKTSYIIMF CKFDEDWD GNLADDWE GSLEDDWD GSLEDDWD	GPDVCGG-SK GPDVCGA-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGFDIK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP	RTHVIMNYKG RVHVILMYKG KVHVIFNYKG KVHVIFNYKG KVHVILMYKG KVHVILHFKN RVHILHYKG LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKP NVSKPADWVD DAAKPEDWDE DAAKPEDWDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENKLIKKEI- KTGVYEEKHA KTGVYEEKHA KTGVFEEKHA : EKEIDDPDDK VKEIDDPEDS RAKIDDPTDS	KC RC RC RC RC RC KRPDADLKTY KRPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEGWDDIP- KPEGWDKP KPEDWDKP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT FTDKKTHLYT FTDKKTHLYT FTDKKTHLYT * * * *
ABCDEFGHIJK ABCDE	CGGGYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLP CGGGYIKVFP CGGGYIKVFP CGGGYIKLP CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS LIRPNNTYV LVISPDNTYE LIVRPDNTYE LIVRPDNTYE	KKSIESEEKF SKLDQSAF SGLDQKDM NSLDQTDM ADADLGDF ADIDQKNE SMDPEKF KTPELNLDQF KTPELNLDQF KTPELNLDQF VKIDNGVEKQE VKIDNSQVES VKIDNSQVES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETEYNVMF HGETEYNVMF HGETEYWLMF HDKTPYTIMF HDKTPYTIMF TDKTSYIIMF YDKTSYIIMF GKFDEDWD GSLEDDWD GSLEDDWD GSLEDDWD	GPDVCGG-SK GPDUCGA-SK GPDICGPCTK GPDICGPCTK GPDICGPCTR GPDICGPCTR GPDICGCPTR GPDKCGEDYK GPDKCGEDYK *** * MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP FLPPKKIKDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILHYNG LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKD DA&SKPEDWDE DASKPEDWDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGVYEEKHA KTGYYEEKHA : EKEIDDPNDK VKEIDDPEDV RAKIDDPTDS RAKIDDPTDS	KC RC RC RF KRPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEGWDDIP- KPEGWDKP KPEDWDKP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDCFTHLYT FTDKKTHLYT FTDKKTHLYT * * **
ABCDEFGHIJK ABCDEF	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLLS CGGAYKLLS CGGAYKLLS CGGAYKLLS LIIRPNNTYV LVISPDNTYK LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNF KTPELNLDQF KTPELNLDQF TDDLILENF DTDDLILENF VKIDQSEKQE VKIDNSQVES VKIDNSQVES VKIDNSQVES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETRYNVMF HDKTPYTIMF HDKTPYTIMF HDKTPYTIMF CKFDEDWD GNLADDWE GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD	GPDVCGG-SK GPDICGPGTK GPDICGPGTK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDICGEDTK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP FLPPKKIKDP FLPPKKIKDP FLPPKKIKDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILNYKG RVHVILHFNYKG HFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKP DAKZPEDWDE DAAKPEDWDE DAKKPEDWDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENKLIKKEI- KTGYYEEKHA KTGYYEEKHA KTGYYEEKHA KTGYYEEKHA KTGYPEKHA KTGDPEDV RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS	KC RC RC RC RF KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEGWDDIP- KPEGWDDIP- KPEDWDKP KPEDWDKP KPEDWDKP KPEDWEKP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KSDELTHLYT PEDKLTHYT FTDKKTHLYT FTDKKTHLYT * :* **
ABCDEFGHIJK ABCDEFG	CGGGYIKLLP CGGSYIKLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYIKLP CGGGYIKLP CGGGYIKLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS LINPDNTYE LINPDNTYE LINPDNTYE LINPDNTYE LINPDNTYE LINPDNTYE LINPDNTYE	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDDLILENF VKIDNGVEKQE VLVDNKEIQA VKIDNSQVES VKIDNSQVES VKIDNSQVES VKIDNSQUES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETRYNVMF HDKTPYTIMF HDKTPYTIMF HDKTPYTIMF CNLADDWE GNLADDWE GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD	GPDVCGG-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGP-TR GPDICGPDIK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK K** ** MLAPKEIDDP LLPSKLIKDP FLPPKKIKDP FLPPKKIKDP LLPAKKIKDP LTSLKKETSP	RTHVIMNYKG RVHVILMYKG KVHVIFNYKG KVHVIFNYKG RVHVILMYKG KVHVILHYKN RVHILHYKG LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKP DAAKPEDWDE DAAKPEDWDE DAAKPEDWDE DAKKPEDWDE DAKKPEDWDE DAKKPEDWDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENKLIKKEI- KTGYYEEKHA KTGYYEEKHA KTGYFEKHA : EKEIDDPDDK RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAVIDAEDA CWEQTKDN	KC RC RC RC RC RC KPPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KAQDWEK	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KDDEFTHLYT KVDGFTHLYT FTDKKTHLYT FTDKKTHLYT * * * **
ABCDEFGHIJK ABCDEFGH	CGGGYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLP CGGGYVKLP CGGGYIKVFP CGGGYIKVFP CGGGYIKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS UIRPNNTYE LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE LINSDNTYE LILSDNTYE LILSDNTYE LILSDNTYE	KKSIESEEKF SKLDQSAF SGLDQKDM NSLDQTDM ADADLGDF ADIDQKNE KTPELNLDQF KTPELNLDQF KTPELNLDQF VKIDNGVEKQE VLVDNKEIQA VKIDNSQVES VKIDNSQVES VKIDNSQVES VKIDNSQVES VKIDNSQVES	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETEYNVMF HGETEYNVMF HDKTPYTIMF HDKTPYTIMF HDKTPYTIMF TDKTSYIIMF YDKTSYIMF GKLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD	GPDVCGG-SK GPDUCGA-SK GPDICGPCTK GPDICGPCTK GPDICGPCTR GPDICGPCTR GPDICGCPTR GPDRCGC-QN GPDKCGEDYK MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP FLPPKKIKDP FLPPKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKEIDDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILNYKG LHFIFRHKNP LHFIFF LHFIFRHKNP LHFIFF	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- ENKLIKKEI- KNVLINKDI- ENKLIKKEI- KTGVYEEKHA KTGYYEEKHA KTGYYEEKHA : EKEIDDPNDK VKEIDDPEDV RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS CHUDPEDK	KVRCRCRF KRPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEOWDRP KPEDWDRP KPEDWDRP KAQDWEK KAQDWEK KPEDWDRE	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDCFTHLYT FTDKKTHLYT FTDKKTHLYT * * **
ABCDEFGHIJK ABCDEFGHI	CGGGYIKLLP CGGSYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS LINPDNTYE LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE LINSDNTYE LILNSDNTYE LILNSDNTYE LILNSDNTYE LILNPDLSFE	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNF KTPELNLDQF KTPELNLDQF TDDLILENF DTDDLILENF VKIDQSVES VKIDNSQVES VKIDNSQVES VKIDGSAQT VKIDGSAQT VKIDGSINS	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF HGETPYNVMF HGETKYWLMF HDKTPYTIMF HDKTPYTIMF CNLADDWE GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GNLCNMTPP	GPDVCGG-SK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDCCGEDYK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK GPDKCGEDYK HLPSKLIKDP FLPPKKIKDP FLPPKKIKDP LLPSKKIKDP LLPSKKIKDP LLPKKIKDP LLPKKIKDP LLPKKIKDP LLPKKIKDP LTSLKKETSP LLPPREIVDE	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG RVHVILNYKG RVHVILNFKN RVHILHFNK LHIFRHKNP LHFIFFH LHFIFH LHFIFH LH	KNNLIRKEI- KNNLIKKEIN KNVLINKDI- KNVLINKDI- ENKLIKKEI- ENKLIKKEI- ENREWSKRI- KTGYYEEKHA KTGYYEEKHA KTGYYEEKHA KTGYYEEKHA KTGYPEKHA KTGDPEDV RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS REYIDDAEDA EETMDDPEDA	KC RC RC RC RF KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPE0WDLP KPE0WDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP KPEDWDKP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KSDELTHLYT FTDKKTHLYT FTDKKTHLYT FTDKKTHLYT * * **
АВСDЕFGHIJК АВСDЕFGHIJ	CGGGYIKLLP CGGSYIKLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLP CGGGYIKLP CGGGYIKLP CGGGYIKLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS CGGAYVKLLS LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE LIVRPDNTYE LILNPDNTYE LILNPDNSFE LILNPDNSFE	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNL SMDPEKF KTPELNLDQF KTPELNLDQF DTDDLILENF VKIDRSVES VKIDNSQVES VKIDNSQVES VKIDNSQVES VKIDRSQESAQT VKIDGSIES FFLDGESKAK ILVDQSIVNS	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGETPYNVMF MGKSQYYIMF HDKTPYTIMF HDKTPYTIMF HDKTPYTIMF CKEDEDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEWF	GPDVCGG-SK GPDICGPGTK GPDICGPGTK GPDICGPGTK GPDICGPOTK GPDICGFDIK GPDRCGEDYK GPDKCGEDYK GPDKCGEDYK K** ** MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP FLPPKKIKDP FLPPKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKEIDP VNPSREIEDP	RTHVIMNYKG RVHVILMYKG KVHVIFNYKG KVHVIFNYKG RVHVILMYKG KVHVILHYKN RVHILHYKN RVHILHFKN RVHILHFKN LHFIFRHKNP LHFIFRHKNP LHFIFRHKP DAKPEDWDE DAKPEDWDE DAKPEDWDE AKKPEDWDE EDRKPEDWDE EDRKPEDWDE	KNNLIRKEI- KNNLIRKEIN KNVLINKDI- ENKLIKKEI- ENRLIKKEI- ENREWSKRI- KTGYYEEKHA KTGYYEEKHA KTGYFEKHA EKEIDDPDK RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPDAEN PFNIPDPEAV	KC RC RC RC RC RC RC KPPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRDAP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDGFTHLYT FTDKKTHLYT FTDKKTHLYT FTDKKTHLYT ETDKKTHLYT AKTPDEEATK AKIPDEEATK
АВСDEFGHIJК АВСDEFGHIJК	CGGGYIKLLP CGGGYVKLFP CGGGYVKLFP CGGGYVKLFP CGGGYVKLP CGGGYKKLP CGGGYKKLP CGGGYKKLP CGGGYKKLS CGGAYKLLS CGGAYKLLS CGGAYKLLS CGGAYKLLS LIRPNNTYV LVISPDNTYK LIVRPDNTYE LINRPDNTYE LILNSDNTYE LILNPDNSFE LILNPDNSFE LILNPDNSFE LINPDNSFE	KKSIESEEKF SKLDQSAF SGLDQKDM MSLDQTDM ADADLGDF ADIDQKNE SMPPEKF KTPELNLDQF KTPELNLDQF KTPELNLDQF VKIDGVEKQE VLVDNKEIQA VKIDNSQVES VKIDNSQVES VKIDGSAQT SKIDQSVNS ILVDQSVNS ILVDQSVNS	TPESEYNIMF DGESEYSIMF HGDSEYNIMF HGDSEYNIMF HGDSEYNIMF HGETRYNMF HGETRYNMF HDKTPYTIMF HDKTPYTIMF HDKTPYTIMF UKTSYIMF CNLADDWE GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD GSLEDWD	GPDVCGG-SK GPDUCGA-SK GPDICGPCTK GPDICGPCTK GPDICGPCTK GPDICGPCTR GPDICGEDTK GPDKCGEDYK GPDKCGEDYK *** ** MLAPKEIDDP LLPSKQIKDP FLPPKKIKDP FLPPKKIKDP FLPPKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKKIKDP LLPAKKIKDP LLPREIDE VNPSREIEDP VNPSREIEDP VNPSREIEDP	RTHVIMNYKG RVHVILNYKG KVHVIFNYKG KVHVIFNYKG RVHVILNYKG RVHVILNYKG KVHVILNYKG LHFIFRHKNP LHFIFRHKNP LHFIFRHKNP LHFIFRHKHP DAKPEDWDE DAKPEDWDE DAKPEDWDE DAKPEDWDE EDKPEDWDE EDRKPEDWDE EDRKPEDWDE	KNNLIRKEI- KNHLIKKEIN KNVLINKDI- ENKLIKKEI- KYHENKKLI- ENREWSKRI- KTGVYEEKHA KTGVYEEKHA KTGVFEEHA EKEIDDPDDK KAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPTDS RAKIDDPEDK RPKIPDPAV RAKIPDPSAV	KCRCRCRF KRPDADLKTY KRPDADLKTY KRPDADLKTY KPPDVDLKKF KPEDWDLP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPEDWDRP KPDWNEDAP KPDDWNEDAP KPDDWNEDAP KPDDWDESEP	ESDDISHLYT ETDQLTHQYT KDDEFTHLYT KDDEFTHLYT KSDELTHLYT KVDCFTHLYT FTDKKTHLYT FTDKKTHLYT * * **

Fig. 2.1 Alignment of Calrs and related proteins from various species. (a) Calr *E. histolytica* XP_655241; (b) Calr *D. discoideum* Q23858; (c) Calr *M. musculus* P14211; (d) Calr *R. norvegicus* P18418; (e) Calr *H. sapiens* P27797; (f) Calr *C. elegans* P27798; (g) Calr3 *H. sapiens* Q96L12; (h) Calr *T. cruzi* XP_812571; (i) Canx *C. lupus* P24643; (j) Canx *H. sapiens* P27824; (k) Clgn *H. sapiens* O14967. The alignment was carried out using Clustal 2.1 and default parameters. Coloring: core region—blue, P-domain—yellow. C-domain—green

C-terminal domain of the molecules, the first constituting the Calr β 19 strand of the upper β -sheet and the second constituting the Calr β 20a/b strand of the lower β -sheet (Fig. 2.2a). From the core extend the proline-rich stretches as β -hairpins (P-domain, Fig. 2.2c), stabilised by hydrogen bonds between four (Canx) or three (Calr) sets of repeat sequences. The α -helix of the globular core is longer in Calr, where it continues into the acidic tail (Fig. 2.2c), while in Canx the α -helix is followed by a transmembrane sequence and the cytoplasmic domain.

A B C D E F G H I J K	K E E E E A PDGWLDDEPE PAGWLDDEPE PAGWLDDEPK	TIVDPNAKKP SIVDPEAVKP HIPDPDAKKP HIPDPDAKKP HIPDPDAKKP HELDASTSKQ MIPDADAKKP YVPDPDAEKP YVPDPDAEKP FIPDPNAEKP	EEWNDEDDGE EDWNEEDDGE EDWDEEMDGE EDWDEEMDGE EDWDEEMDGE EDWDDEMDGE SDWNGDLDGD DDWDDAEDGP EDWDEMDGE EDWDEMDGE EDWDEMDGE	WEAPTIE WEAPTIA WEPPVIQ WEPPVIQ WEPPMID WEAPMID WEAPMID WEAPMIAP WEAPQIANPK WEAPQIANPR WEAPQILNPA	CESAPGCGVW CESAPGCGVW CRIGCGEW	NPE NPE NPE NPE NPE NPE NPE NPK (QRPMIDNPN) (RPVIDNPN)	<pre>KGEWKPKRIF KGEWKAKKIF KGEWKPRQID KGEWKPRQID KGEWKPRQID KGEWKPRQIF KGEWKPRQIF KGKWKPPMID KGKWKPPMID KGWWRPPLVD</pre>	NPAYKGEWVH NPEYKGEWVH NPDYKGTWIH NPDYKGTWIH NPAYKGKWIH -GIHKDVWLH NPAYKGVWEP NPAYKGVWEP NPNYQGIWKP NPSYQGIWKP NPNYQGIWKP
	DOTANDOVAN				TTODICCAC			
ABCDEFGHIJK	PUIDNPEYAE PEIDNPEYSP PEIDNPEYSP PEIDNPEYSP PEIENPEYTP RKMKNTDYLT RKIPNPDFFE RKIPNPDFFE RKIPNPDFFE RKIPNPDYFE	DPELYKYDS- DNELYLFND- DANIYAYDS- DANIYAYDS- DPSIYAYDN- DDELYSYES- QYDLSEFEN- DLEPFKMTP- DLEPFKMTP- DLEPFRMTP- DLPFFRTP-	FAVLGLDUWQ FAVLGLDLWQ FAVLGLDLWQ FGVLGLDLWQ WGAIGFDLWQ WGAIGELWQ PLTHVGIDVWQ FSAIGLELWS FSAIGLELWS	VKAGIITDDI VKSGIFNNM VKSGIFDNF VKSGTIFDNF VKSGTIFDNF VRSGTIFDNI VRSGIFDNF MTSDIFFDNF MTSDIFFDNF MTSDIFFDNF	LITDDLEEAE IVTDSVEEAK LITNDEAYAE LITNDEAYAE LITNDEAYAE IITDSVEEAE LITDDEEYAD MIGDDLKEVL IVCGDRRVVD IICADRRIVD IICSEKEVAD	KEAKVILEK- DFSEKTFVAN EFGNETWGVT EFGNETWGVT AHAAETFDKL NFGKATWGET ELVEKTYGSL DWANDGWGLK HWAADGWRWK	QEAE K K K K K	VVGQMIEAAE VVGQMIEAAE VVGQMIEAAE
	: *.::	:	* *	: :. :.::	::			
A B C D E F G H I J K	ERPWLWVVVV GHPWLWLIYL	LTVALPVFLV LTVALPVFLV VTAGVPIALI	/ ILFCCSGK / ILFCCSGK / ILFCCSGK TSFCWPRKVK	KQSSPVEYKK KQTSGMEYKK	K	AI 	AAEKKMRDEI AAEKQMKDKQ AAEKQMKDKQ AAEKQMKDKQ AAEKQMKDKQ TVEKEKKEKA GPEREMDAIQ ADALKVMEDM EEKDKGDEE AALEKPMDLE	KEAEKQKEEE LAAEKAAEKE DEEQRLKEEE DEEQRLKEEE DEEQRLKEEE DEETRKAEEE AKEEMKKARE EKEKRKKEEE EEGEEKLEEK EEGEEKLEEK
A B C D E F G H I J	AKKEAEKQKE AEEADEEEEE EDKKRKEEEE EDKKRKEEEE ARKKAEEEK EEEELLSGK EEKEKAEE QKSDAEEDGG QKSDAEEDGG	EETKEEIKKE VAEEDLVKTD AEDKEDDDR AEDKEDEDDR AEDKEDDEDK AKKDDDEEEK INRHEHYFNQ EKDEEELEEK TASQEEDDRK TVSQEEEDRK	ENKEEL DKKEEVKKST DEDEDEDEK DEDEDEDEK DEDEDEEDEK EEEGHDEL- FHRRNEL GDGDKEDL PKAEEDEILN PKAEDDEILN	KKVDHDEL EEDEE-ESPG EEDEE-DATG EEDEEEDVPG 	QAKDEL QAKDEL QAKDEL QAKDEL E		389 424 416 416 317 384 401 593 592	

Fig. 2.1 (continued)

There are several interesting characteristic features of the conserved structure of Calr. A carbohydrate-binding site on the upper concave β -sheet capable of accommodating a tetrasaccharide and having affinity for terminal glucose residues. This site contains several conserved residues (cluster 1) of importance for carbohydrate binding (Kozlov et al. 2010a, b; Chouquet et al. 2011; Moreau et al. 2016).

An exposed disulphide bridge ($C^{106}-C^{137}$ in hCalr) is located on the edge of the carbohydrate-binding site. This bridge is unusual due to its surface location and its geometry, which reveals an almost parallel, strained conformation with a sulphur atom distance of 2.8–3 Å (in contrast to normal unstrained disulphide bonds, which have a torsion angle of approximately 90° and a sulphur atom distance of 2 Å (Thornton 1981). The disulphide bridge is important for thermodynamic stability

Protein (fragment)	Method	Comments	References
rCalr	NMRS	P hairpin	Ellgaard et al. (2001a, b, 2002)
mCalr	XRC	18-206, 301-368, C163S	Kozlov et al. (2010a)
mCalr	XRC	Globular core, C163S	Pocanschi et al. (2011)
hCalr	XRC	Globular core	Chouquet et al. (2011)
hCalr	XL-MS	Native	Boelt et al. (2016)
tcCalr	XRC, SAXS	Globular core	Moreau et al. (2016)
ehCalr	XRC	Globular core	Moreau et al. (2016)
hCalr	ECM	Complex with MHC	Blees et al. (2017)
Crt C-domain	CDS, MM	C-"tail", Ca ²⁺ sensor	Villamil Giraldo et al. (2010)
hCalr	XRC	Complex with GABARAP	Thielmann et al. (2009)
hCalr	SAXS	Native	Nørgaard Toft et al. (2008)
m/hCalr, Canx, Clgn	XRC, NMR	P hairpin cyclophilin/ERp29 complex	Kozlov et al. 2010b, (2017)
dCanx	XRC	Fragment (47-468)	Schrag et al. (2001), Hahn et al. (1998)
dCanx	NMRS	P hairpin tip	Pollock et al. (2004)

Table 2.1 Structural data available for Calr and Canx

CDS circular dichroism spectroscopy, *d* dog, *ECM* electron cryo-microscopy, *eh E. histolytica, h* human, *m* mouse, *MM* molecular modelling, *NMRS* nuclear magnetic resonance spectroscopy, *r* rat, *tc T. cruci, XL-MS* cross-linking/mass spectrometry, *XRC* X-ray crystallography

and reduceable by DTT but is not affected by alkylating agents under conditions, where the free Cys¹⁶³ in hCalr can be alkylated (100 mM sodium phosphate, pH 7.2, room temperature), showing that it is stable under near-to-physiological conditions (Houen and Koch 1994; Højrup et al. 2001; Jørgensen et al. 2005), in accordance with its surface-exposed localisation at the edge of the carbohydrate-binding site. Interestingly, the disulphide bridge seems to be important for lectin activity (Kozlov et al. 2010a).

The lower β -sheet has a conserved structural Ca²⁺ ion-binding site (Fig. 2.2a) and also contains several residues conserved in published structures of Calr and Caln (cluster 2) and in primary sequences of Calr and related proteins (Fig. 2.1), including a triad composed of D¹⁶⁶, H¹⁷⁰ and D¹⁸⁷ (DHD triad). Finally, the C-terminal α -helix contains two conserved residues including W³⁴⁷, close to the DHD triad (Fig. 2.2d) (Kozlov et al. 2010a, b; Chouquet et al. 2011; Moreau et al. 2016).

Chemical cross-linking in solution in combination with mass spectrometry (MS) of human Calr has confirmed the overall structure of the globular domain and has added some structural information on the parts of Calr absent in the crystallised "amputated" Calr (Fig. 2.2b). In the presence of Ca^{2+} , the P-domain hairpin folds back on itself and associates with the C-domain, which itself forms an α -helix extending "upwards" to meet the P-domain. Together, they cover the lectin site, however, in the absence of Ca^{2+} , they are more flexible and do not sterically



Fig. 2.2 *H. Sapiens* (Hs) Calr structures. (a) XRC core structure (Chouquet et al. 2011). The structural Ca^{2+} ion is shown in yellow. (b) XL-MS structure (Boelt et al. 2016). (c) HsCalr (red) in complex with MHC-I (yellow), Pdia3 (blue), tapasin (cyan) and β 2-microglobulin (green) (Blees et al. 2017). (d) Zoom on the cluster2 residues common to Calr (cyan) and Canx (green) proteins

limit access to the lectin site to the same extent. X-ray studies of *Trypanosoma cruzi* and *Entamoeba histolytica* Calrs have also shown for the first time a possible hinge motion at the basis of the Calr P-domain, as a switch between an open and a more compact conformation. This part of the Calr molecule differs in Canx (Moreau et al. 2016).

Structural studies of full-length Calr proteins have been hindered by its flexibility. However, its overall ordered structure could be trapped by cryo-electron microscopy in the context of the MHC-I peptide-loading complex (Blees et al. 2017). In this complex, Calr still shows substantial flexibility, as compared to its structurally stable binding partners. This flexibility remains despite the fact that it interacts with an MHC-class I glycan through its carbohydrate-binding site, with Pdia3 through the tip of its P-arm and with tapasin through its acidic tail.

The overall structure of Calr and Canx indicated by the structural data for various amputated forms of the molecules has been confirmed by several studies with other methods including electron cryo-microscopy (ECM), small angle X-ray scattering (SAXS) and nuclear magnetic resonance (NMR) spectroscopy (Table 2.1). Both molecules have a stable globular beta sandwich core with an α helix at the bottom.

From the core protrudes a β -hairpin P-domain and a C-domain, which in the presence of Ca²⁺ make contacts and cover the lectin site.

2.4 Calr Structure-Function Relationships

The thermodynamic stability of Calr is moderate and depends on the general packing of the core region, the conserved, surface-exposed disulphide bridge and Ca²⁺ association (Jørgensen et al. 2005; Duus et al. 2013). Calr and Canx are chaperones for glycoprotein synthesis and several of the conserved residues in cluster 1, including the disulphide bridge, have been shown to be involved in lectin activity of Calr/ Canx (Kozlov et al. 2010a, b). These residues are also strikingly conserved in the parasite Calrs (*T. cruzi, E. histolytica*) (Moreau et al. 2016). A putative polypeptide binding site has been located at the edge of the carbohydrate-binding upper β -sheet and Calr has been shown to interact with several non-glycosylated proteins (Duus et al. 2009; Pocanschi et al. 2011; Chouquet et al. 2011; Møllegaard et al. 2011; Moreau et al. 2016).

Calr has been implicated in several diseases including cancers, autoimmune diseases, and others (Tesniere et al. 2008; Gold et al. 2010; Zamanian et al. 2013; Wiersma et al. 2015; Eggleton et al. 2016; Schcolnik-Cabrera et al. 2019; Houen 2019). Most evident in relation to structure are somatic alterations seen in myeloproliferative neoplasms (MPNs), including polycythemia vera, myelofibrosis and essential thrombocythemia (Klampfl et al. 2013; Nangalia et al. 2013; Pietra et al. 2016; Imai et al. 2017). These alterations are mainly insertions and deletions (INDELS) in exon 9 leading to a common polybasic stretch of amino acids in the C-terminus, in close conjunction to the α -helix in the C-domain. Since these INDELS have only been detected in MPNs, it can be speculated that the frameshifted Calr forms allow the affected myeloid precursor cells to remain viable, to grow and differentiate to some degree and to exhibit neoplastic properties. The structural polybasic alterations of the C-terminus can be speculated to mimic the binding of Ca²⁺ to the acidic C-terminus of wild type Calr, allowing the protein to remain stable and viable, but imparting some new properties on the molecule (i.e. an oncogenic alteration). In agreement with this, the frameshifted Calr has been shown to interact with the thrombopoietin receptor, which itself can cause MPN upon mutation and which signals through a pathway involving the Janus kinase (JAK), mutations of which is the most prominent cause of MPN (Vainchenker et al. 2019).

In relation to cancer in general, Calr has been shown to translocate to the cell surface in response to various forms of cellular stress, where it may serve as a signal for immune recognition (Tesniere et al. 2008; Wiersma et al. 2015; Eggleton et al. 2016; Schcolnik-Cabrera et al. 2019). However, it is presently unclear, whether the various involvements of Calr in cell transformation and immune surveillance are related to structural alterations or to changes in concentration and subcellular localization.

2.5 Summary Points

- Calr and Canx are conserved Ca²⁺-binding molecular chaperones in eucaryotes.
- Both Calr and Canx have a globular β -sandwich/ α -helix core with a structural Ca²⁺ ion. From the core extends a characteristic β -hairpin P-domain and a C-domain. In Calr, the C-domain is acidic and has a large capacity for Ca²⁺ binding. In Canx, the C-domain has a transmembrane stretch and a cytoplasmic domain.
- The upper β -sheet of the globular core sandwich has a carbohydrate-binding site and a disulphide bridge at the edge, which both are involved in carbohydrate interactions. A polypeptide binding site is located at the edge of the carbohydratebinding site but has not been precisely mapped yet.
- The lower β -sheet of the globular sandwich has a conserved D¹⁶⁶-H¹⁷⁰-D¹⁸⁷ triad, a free C¹⁶³ and several other conserved residues, including Y¹⁷² (hCalr numbering) (Fig. 2.2d).
- The C-terminal α -helix at the bottom of the α -helix core has a conserved W³⁴⁷ residue facing the D¹⁶⁶-H¹⁷⁰-D¹⁸⁷ triad (hCalr numbering) (Fig. 2.2d).
- The structure of Calr/Canx proteins reflects their multiple functions and interaction partners, including nascent (glycosylated) proteins, cyclophilin, Pdia3, and others, as illustrated by the role of Calr and Canx in the peptide loading complex.
- Characteristic mutations in the exon coding for the extreme C-terminus of hCalr changes the polyacidic Ca²⁺-binding sequence to a polybasic sequence.

2.6 Future Issues

- Detailed structural analysis of complete Calr/Canx proteins in complex with substrates and binding partners.
- Detailed analysis of the role of individual amino acids in the thermodynamic stability and functions of Calr and Canx.
- Detailed understanding of the role of the polybasic sequence in frameshifted Calr variants involved in myeloproliferative neoplasms.
- Detailed understanding of cell surface functions of Calr.

Conflicts-of-Interest Disclosure Statement GH, PHP, EC, CG, RS declare no conflicts-of-interest.

References

Bergeron JJ, Brenner MB, Thomas DY, Williams DB (1994) Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. Trends Biochem Sci 19:124–128. https://doi.org/10. 1016/0968-0004(94)90205-4

- Blees A, Januliene D, Hofmann T, Koller N, Schmidt C, Trowitzsch S, Moeller A, Tampé R (2017) Structure of the human MHC-I peptide-loading complex. Nature 551:525–528. https://doi.org/ 10.1038/nature24627
- Boelt SG, Norn C, Rasmussen MI, André I, Čiplys E, Slibinskas R, Houen G, Højrup P (2016) Mapping the Ca²⁺ induced structural change in calreticulin. J Proteome 142:138–148. https:// doi.org/10.1016/j.jprot.2016.05.015
- Chouquet A, Païdassi H, Ling WL, Frachet P, Houen G, Arlaud GJ, Gaboriaud C (2011) X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multi-molecular mechanism. PLoS One 6:e17886. https://doi.org/10.1371/journal.pone. 0017886
- Coe H, Michalak M (2010) ERp57, a multifunctional endoplasmic reticulum resident oxidoreductase. Int J Biochem Cell Biol 42:796–799. https://doi.org/10.1016/j.biocel.2010.01.009
- Danilczyk UG, Cohen-Doyle MF, Williams DB (2000) Functional relationship between calreticulin, calnexin, and the endoplasmic reticulum luminal domain of calnexin. J Biol Chem 275:13089–13097. https://doi.org/10.1074/jbc.275.17.13089
- de Virgilio C, Bürckert N, Neuhaus JM, Boller T, Wiemken A (1993) CNE1, a Saccharomyces cerevisiae homologue of the genes encoding mammalian calnexin and calreticulin. Yeast 9:185–188. https://doi.org/10.1002/yea.320090209
- Del Bem LE (2011) The evolutionary history of calreticulin and calnexin genes in green plants. Genetica 139:255–259. https://doi.org/10.1007/s10709-010-9544-y
- Duus K, Sandhu N, Jørgensen CS, Hansen PR, Steinø A, Thaysen-Andersen M, Højrup P, Houen G (2009) Interaction of the chaperone calreticulin with proteins and peptides of different structural classes. Protein Pept Lett 16:1414–1423. https://doi.org/10.2174/092986609789353772
- Duus K, Larsen N, Tran TA, Güven E, Skov LK, Jespersgaard C, Gajhede M, Houen G (2013) Chemical and thermal unfolding of calreticulin. Protein Pept Lett 20:562–568. https://doi.org/ 10.2174/0929866511320050009
- Eggleton P, Bremer E, Dudek E, Michalak M (2016) Calreticulin, a therapeutic target? Expert Opin Ther Targets 20:1137–1147. https://doi.org/10.1517/14728222.2016.1164695
- Ellgaard L, Riek R, Herrmann T, Güntert P, Braun D, Helenius A, Wüthrich K (2001a) NMR structure of the calreticulin P-domain. Proc Natl Acad Sci U S A 98:3133–3138. https://doi.org/ 10.1073/pnas.051630098
- Ellgaard L, Riek R, Braun D, Herrmann T, Helenius A, Wüthrich K (2001b) Three-dimensional structure topology of the calreticulin P-domain based on NMR assignment. FEBS Lett 488:69–73. https://doi.org/10.1016/s0014-5793(00)02382-6
- Ellgaard L, Bettendorff P, Braun D, Herrmann T, Fiorito F, Jelesarov I, Güntert P, Helenius A, Wüthrich K (2002) NMR structures of 36 and 73-residue fragments of the calreticulin P-domain. J Mol Biol 322:773–784. https://doi.org/10.1016/s0022-2836(02)00812-4
- Gold LI, Eggleton P, Sweetwyne MT, Van Duyn LB, Greives MR, Naylor SM, Michalak M, Murphy-Ullrich JE (2010) Calreticulin: non-endoplasmic reticulum functions in physiology and disease. FASEB J 24:665–683. https://doi.org/10.1096/fj.09-145482
- Hahn M, Borisova S, Schrag JD, Tessier DC, Zapun A, Tom R, Kamen AA, Bergeron JJ, Thomas DY, Cygler M (1998) Identification and crystallization of a protease-resistant core of calnexin that retains biological activity. J Struct Biol 123:260–264. https://doi.org/10.1006/jsbi.1998. 4032
- Halperin L, Jung J, Michalak M (2014) The many functions of the endoplasmic reticulum chaperones and folding enzymes. IUBMB Life 66:318–326. https://doi.org/10.1002/iub.1272
- Højrup P, Roepstorff P, Houen G (2001) Human placental calreticulin characterization of domain structure and post-translational modifications. Eur J Biochem 268:2558–2565. https://doi.org/ 10.1046/j.1432-1327.2001.02138.x
- Houen G (2019) COMMENTARY: calreticulin oncogene, anti-oncogene, or both? Curr Protein Pept Sci 20:111–112. https://doi.org/10.2174/138920372001181031111149

- Houen G, Koch C (1994) Human placental calreticulin: purification, characterization and association with other proteins. Acta Chem Scand 48:905–911. https://doi.org/10.3891/acta.chem. scand.48-0905
- Ikawa M, Tokuhiro K, Yamaguchi R, Benham AM, Tamura T, Wada I, Satouh Y, Inoue N, Okabe M (2011) Calsperin is a testis-specific chaperone required for sperm fertility. J Biol Chem 286:5639–5646. https://doi.org/10.1074/jbc.M110.140152
- Imai M, Araki M, Komatsu N (2017) Somatic mutations of calreticulin in myeloproliferative neoplasms. Int J Hematol 105:743–747. https://doi.org/10.1007/s12185-017-2246-9
- Jia XY, He LH, Jing RL, Li RZ (2009) Calreticulin: conserved protein and diverse functions in plants. Physiol Plant 136:127–138. https://doi.org/10.1111/j.1399-3054.2009.1223.x
- Jørgensen CS, Trandum C, Larsen N, Ryder LR, Gajhede M, Skov LK, Højrup P, Barkholt V, Houen G (2005) Conformational stability of calreticulin. Protein Pept Lett 12:687–693. https:// doi.org/10.2174/0929866054696082
- Kepp O, Galluzzi L (eds) (2020) Biology of the endoplasmic reticulum. Int Rev Cell Mol Biol 350:1–325. https://doi.org/10.1016/S1937-6448(20)30020-4
- Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NC, Berg T, Gisslinger B, Pietra D, Chen D, Vladimer GI, Bagienski K, Milanesi C, Casetti IC, Sant'Antonio E, Ferretti V, Elena C, Schischlik F, Cleary C, Six M, Schalling M, Schönegger A, Bock C, Malcovati L, Pascutto C, Superti-Furga G, Cazzola M, Kralovics R (2013) Somatic mutations of calreticulin in myeloproliferative neoplasms. N Engl J Med 369:2379–2390. https://doi.org/10.1056/NEJMoa1311347
- Kozlov G, Pocanschi CL, Rosenauer A, Bastos-Aristizabal S, Gorelik A, Williams DB, Gehring K (2010a) Structural basis of carbohydrate recognition by calreticulin. J Biol Chem 285:38612–38620. https://doi.org/10.1074/jbc.M110.168294
- Kozlov G, Bastos-Aristizabal S, Määttänen P, Rosenauer A, Zheng F, Killikelly A, Trempe JF, Thomas DY, Gehring K (2010b) Structural basis of cyclophilin B binding by the calnexin/ calreticulin P-domain. J Biol Chem 285:35551–35557. https://doi.org/10.1074/jbc.M110. 160101
- Kozlov G, Muñoz-Escobar J, Castro K, Gehring K (2017) Mapping the ER interactome: the P domains of calnexin and calreticulin as plurivalent adapters for foldases and chaperones. Structure 25:1415–1422. https://doi.org/10.1016/j.str.2017.07.010
- Matsusaki M, Kanemura S, Kinoshita M, Lee YH, Inaba K, Okumura M (2020) The protein disulfide isomerase family: from proteostasis to pathogenesis. Biochim Biophys Acta 1864:129338. https://doi.org/10.1016/j.bbagen.2019.04.003
- McCauliffe DP, Yang YS, Wilson J, Sontheimer RD, Capra JD (1992) The 5'-flanking region of the human calreticulin gene shares homology with the human GRP78, GRP94, and protein disulfide isomerase promoters. J Biol Chem 267:2557–2562
- Michalak M, Mariani P, Opas M (1998) Calreticulin, a multifunctional Ca²⁺ binding chaperone of the endoplasmic reticulum. Biochem Cell Biol 76:779–785. https://doi.org/10.1139/bcb-76-5-779
- Møllegaard KM, Duus K, Træholt SD, Thaysen-Andersen M, Liu Y, Palma AS, Feizi T, Hansen PR, Højrup P, Houen G (2011) The interactions of calreticulin with immunoglobulin G and immunoglobulin Y. Biochim Biophys Acta 1814:889–899. https://doi.org/10.1016/j.bbapap. 2011.03.015
- Moreau C, Cioci G, Iannello M, Laffly E, Chouquet A, Ferreira A, Thielens NM, Gaboriaud C (2016) Structures of parasite calreticulins provide insights into their flexibility and dual carbohydrate/peptide-binding properties. IUCrJ 3:408–419. https://doi.org/10.1107/ S2052252516012847
- Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, Avezov E, Li J, Kollmann K, Kent DG, Aziz A, Godfrey AL, Hinton J, Martincorena I, Van Loo P, Jones AV, Guglielmelli P, Tarpey P, Harding HP, Fitzpatrick JD, Goudie CT, Ortmann CA, Loughran SJ, Raine K, Jones DR, Butler AP, Teague JW, O'Meara S, McLaren S, Bianchi M, Silber Y, Dimitropoulou D, Bloxham D, Mudie L, Maddison M, Robinson B, Keohane C, Maclean C, Hill K, Orchard K,

Tauro S, Du MQ, Greaves M, Bowen D, BJP H, Harrison CN, NCP C, Ron D, Vannucchi AM, Papaemmanuil E, Campbell PJ, Green AR (2013) Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N Engl J Med 369:2391–2405. https://doi.org/10. 1056/NEJMoa1312542

- Nørgaard Toft K, Larsen N, Steen Jørgensen F, Højrup P, Houen G, Vestergaard B (2008) Small angle X-ray scattering study of calreticulin reveals conformational plasticity. Biochim Biophys Acta 1784:1265–1270. https://doi.org/10.1016/j.bbapap.2008.05.005
- Ohsako S, Hayashi Y, Bunick D (1994) Molecular cloning and sequencing of calnexin-t. An abundant male germ cell-specific calcium-binding protein of the endoplasmic reticulum. J Biol Chem 269:14140–14148
- Parlati F, Dignard D, Bergeron JJ, Thomas DY (1995) The calnexin homologue cnx1+ in Schizosaccharomyces pombe, is an essential gene which can be complemented by its soluble ER domain. EMBO J 14:3064–3072
- Pietra D, Rumi E, Ferretti VV, Di Buduo CA, Milanesi C, Cavalloni C, Sant'Antonio E, Abbonante V, Moccia F, Casetti IC, Bellini M, Renna MC, Roncoroni E, Fugazza E, Astori C, Boveri E, Rosti V, Barosi G, Balduini A, Cazzola M (2016) Differential clinical effects of different mutation subtypes in CALR-mutant myeloproliferative neoplasms. Leukemia 30:431–438. https://doi.org/10.1038/leu.2015.277
- Pocanschi CL, Kozlov G, Brockmeier U, Brockmeier A, Williams DB, Gehring K (2011) Structural and functional relationships between the lectin and arm domains of calreticulin. J Biol Chem 286:27266–27277. https://doi.org/10.1074/jbc.M111.258467
- Pollock S, Kozlov G, Pelletier MF, Trempe JF, Jansen G, Sitnikov D, Bergeron JJ, Gehring K, Ekiel I, Thomas DY (2004) Specific interaction of ERp57 and calnexin determined by NMR spectroscopy and an ER two-hybrid system. EMBO J 23:1020–1029. https://doi.org/10.1038/sj. emboj.7600119
- Schcolnik-Cabrera A, Oldak B, Juárez M, Cruz-Rivera M, Flisser A, Mendlovic F (2019) Calreticulin in phagocytosis and cancer: opposite roles in immune response outcomes. Apoptosis 24:245–255. https://doi.org/10.1007/s10495-019-01532-0
- Schrag JD, Bergeron JJ, Li Y, Borisova S, Hahn M, Thomas DY, Cygler M (2001) The structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol Cell 8:633–446. https://doi.org/10.1016/s1097-2765(01)00318-5
- Tesniere A, Panaretakis T, Kepp O, Apetoh L, Ghiringhelli F, Zitvogel L, Kroemer G (2008) Molecular characteristics of immunogenic cancer cell death. Cell Death Differ 15:3–12. https:// doi.org/10.1038/sj.cdd.4402269
- Thielmann Y, Weiergräber OH, Mohrlüder J, Willbold D (2009) Structural framework of the GABARAP-calreticulin interface--implications for substrate binding to endoplasmic reticulum chaperones. FEBS J 276:1140–1152. https://doi.org/10.1111/j.1742-4658.2008.06857.x
- Thornton JM (1981) Disulphide bridges in globular proteins. J Mol Biol 151:261-287
- Tjoelker LW, Seyfried CE, Eddy RL Jr, Byers MG, Shows TB, Calderon J, Schreiber RB, Gray PW (1994) Human, mouse, and rat calnexin cDNA cloning: identification of potential calcium binding motifs and gene localization to human chromosome 5. Biochemistry 33:3229–3236. https://doi.org/10.1021/bi00177a013
- Vainchenker W, Plo I, Marty C, Varghese LN, Constantinescu SN (2019) The role of the thrombopoietin receptor MPL in myeloproliferative neoplasms: recent findings and potential therapeutic applications. Expert Rev Hematol 12:437–448. https://doi.org/10.1080/17474086. 2019.1617129
- Villamil Giraldo AM, Lopez Medus M, Gonzalez Lebrero M, Pagano RS, Labriola CA, Landolfo L, Delfino JM, Parodi AJ, Caramelo JJ (2010) The structure of calreticulin C-terminal domain is modulated by physiological variations of calcium concentration. J Biol Chem 285:4544–4553. https://doi.org/10.1074/jbc.M109.034512
- Vincenz-Donnelly L, Hipp MS (2017) The endoplasmic reticulum: a hub of protein quality control in health and disease. Free Radic Biol Med 108:383–393. https://doi.org/10.1016/j. freeradbiomed.2017.03.031

- Wada I, Imai S, Kai M, Sakane F, Kanoh H (1995) Chaperone function of calreticulin when expressed in the endoplasmic reticulum as the membrane-anchored and soluble forms. J Biol Chem 270:20298–20304. https://doi.org/10.1074/jbc.270.35.20298
- Wasąg P, Grajkowski T, Suwińska A, Lenartowska M, Lenartowski R (2019) Phylogenetic analysis of plant calreticulin homologs. Mol Phylogenet Evol 134:99–110. https://doi.org/10.1016/j. ympev.2019.01.014
- Watanabe D, Yamada K, Nishina Y, Tajima Y, Koshimizu U, Nagata A, Nishimune Y (1994) Molecular cloning of a novel Ca²⁺-binding protein (calmegin) specifically expressed during male meiotic germ cell development. J Biol Chem 269:7744–7749
- Wiersma VR, Michalak M, Abdullah TM, Bremer E, Eggleton P (2015) Mechanisms of translocation of ER chaperones to the cell surface and immunomodulatory roles in cancer and autoimmunity. Front Oncol 5:7. https://doi.org/10.3389/fonc.2015.00007
- Zamanian M, Veerakumarasivam A, Abdullah S, Rosli R (2013) Calreticulin and cancer. Pathol Oncol Res 19:149–154. https://doi.org/10.1007/s12253-012-9600-2