



## Yeast Sequencing Reports

# *CNE1*, a *Saccharomyces cerevisiae* Homologue of the Genes Encoding Mammalian Calnexin and Calreticulin

CLAUDIO DE VIRGILIO, NIELS BÜRCKERT, JEAN-MARC NEUHAUS, THOMAS BOLLER AND ANDRES WIEMKEN\*

*Department of Botany, University of Basel, Hebelstr. 1, CH-4056 Basel, Switzerland*

Received 14 July 1992; accepted 6 September 1992

KEY WORDS — *Saccharomyces cerevisiae*; chromosome I; calnexin homologue; *CNE1*.

A DNA fragment of *Saccharomyces cerevisiae* with high homology to the calreticulin genes of different species (Fliegel *et al.*, 1989; Smith and Koch, 1989; Murphy *et al.*, 1990; Rokeach *et al.*, 1991) and the calnexin gene of *Canis domesticus* (Wada *et al.*, 1991) was cloned serendipitously from a genomic DNA library (Rose *et al.*, 1987). It contains an open reading frame of 1506 nucleotides encoding a predicted gene product of 56.9 kDa. In addition, 258 nucleotides at the 5' non-coding region and 264 nucleotides at the 3' non-coding region were sequenced (Figure 1). Pulse-field electrophoresis of yeast chromosomal DNA followed by Southern blot analysis showed the gene to be localized on chromosome I. Furthermore, based on an analysis of restriction sites, the gene could be assigned to the cloned insert  $\lambda$ R23d (Kaback, 1992, personal communication; Kaback *et al.*, 1989).

Computer analysis of the deduced amino acid sequence (Devereux *et al.*, 1984) revealed that the *S. cerevisiae* gene product shares significant similarity with the amino acid sequences of the calreticulins of *Homo sapiens* (48.8%), *Rattus norvegicus* (48.9%), *Mus musculus* (48.9%), *Oryctolagus* sp. (51.4%) and *Onchocerca volvulus* (54.4%), respectively (Figure 2).

The amino-terminal sequence of the *S. cerevisiae* gene product matches the features of a eukaryotic signal peptide which is sufficient to promote the translocation of a polypeptide across the endoplasmic reticulum (ER) membrane (von Heijne, 1990). However, the *S. cerevisiae* sequence lacks the C-terminal ER retention signal KDEL (Pelham, 1990) which is common to all mammalian calreticulins. In addition, the *S. cerevisiae* sequence shares an even higher degree of similarity (56.1%) with the sequence of the *C. domesticus* calnexin (Wada *et al.*, 1991). Calnexin has been suggested to be a trans-membrane protein of the ER membrane with a typical trans-membrane sequence and a cytosolic C-terminus that can be phosphorylated (Wada *et al.*, 1991). The *S. cerevisiae* sequence also has a hydrophobic stretch in the C-terminal region which might function as a trans-membrane domain. The putative cytosolic tail is very short when compared with that of *C. domesticus*. Taken together, we suggest that the presented *S. cerevisiae* gene is the homologue of the *C. domesticus* calnexin gene and we propose the designation: *CNE1*. The complete DNA sequence of *CNE1* has been deposited at the EMBL database (accession no. X66470).

\*To whom correspondence should be addressed.

1 TTTCTATCTCTTCATTCTTTGTACATTAGATACATGGTTTTAGCTTAGTTTTATTTATTTTTATATATCTGGATGTACTATTATTGAAAACTTC  
101 ATTAATAGTTACAACTTTTCAATATCAAGTTGATTAAGAAAAAGAAAATTATTATGGGTTAGCTGAAAACCGTGTGATGCATGTCGTTTAAAGGATTGTG  
201 TAAAAAGTGAACGGCAACGCATTCTAATATAGATAACGGCCACACAAAGTAGTACTATGAAATTTCTGCGTATTATGGTGGCTGTTTTGAATCTA  
1 M K F S A Y L W W L F L N L  
301 GCGTTGGTGAAGGCACTTCATTGCTATCCAACGTTACATTAGCGGAAGATTCTTTCTGGGAGCATTTTCAGGCTTACACTAATACAAAGCATTAAACC  
15 A L V K G T S L L S N V T L A E D S F W E H F Q A Y T N T K H L N Q  
401 AAGAGTGGATCACAAGTGAAGCCGTCAACAATGAAGGCTCTAAAATATATGGTGACAATGGCGACTATCACAGGTCGATGCAAGGATCCGCATGGGA  
49 E W I T S E A V N N E G S K I Y G A Q W R L S Q G R L Q G S A W D  
501 TAAAGGAATCGCAGTTCGAACAGGCAATGCCGAGCTATGATAGGACATCTCTGGAGACACCTATTAATGTTTCAGAAACGGATACCTGGTTGTCCAG  
82 K G I A V R T G N A A A M I G H L L E T P I N V S E T D T L V V Q  
601 TACGAAATTAAGTTGGACAATTCTTTGACGTGCGCGGTCGCTTTTAAAGTTAATGCTGCTGTTTCATGAATGTTGAAGCATTAAAACACTATGCACCCG  
115 Y E I K L D N S L T C G G A F I K L M S G F M N V E A L K H Y A P D  
701 ATACAGAGGGTGTGCGATTAGTTTTGGTCCGGATTATTGTGCTCCTGAAATAAATGGCGTCAATTGCCCATAAAGTTGACAAGATCACACATGA  
149 T E G V E L V F G P D Y C A P E I N G V Q F A I N K V D K I T H E  
801 ATCTAAACTAAGATATTGCAAGAGATGCCCTGTCAAATTAAGTATGATACCTCGCAATCTCATCTGTATACGCTCATAATAGATGAATCAGCGAGTCT  
182 S K L R Y L Q E M P L S K L T D T S Q S H L Y T L I I D E S A Q S  
901 TTTCAAATCTTATAGACGGTAAGACGGTTATGGTAAGAGAACATATCGAAGACAAGAAAAGGTC AATTTTGGCCACCCATTACACCGCCTTAAATGA  
216 F Q I L I D G K T V M V R E H I E D K K K V N F E P P I T P P L M I  
1001 TTCCTGATGTTTCAGTAGCGAAAACCGCATGATTGGGATGATCGCATCCGAATCCAGATCCTGAGGCGGTGAAGCTCAGTGTGCGGATGAACGAGACC  
249 P D V S V A K P H D W D D R I R I P D P E A V K L S D R D E R D P  
1101 ATTGATGATCCACATCCAGATGGCACTGAACCACCAGAATGGAACAGCTCCATCCCGAATACATCTTGACCCAAATGCTCAAAAGCCCTCGTGGTGG  
282 L M I P H P D G T E P P E W N S S I P E Y I L D P N A Q K P S W W  
1201 AAGGAACCTGAGCACGGGAATGGATACCGCCCATGATTAATACTCTTTGCACTGCAGAACCTGGTTGTGGCCAGAGATACCAGGCGTATAAATA  
316 K E L E H G E W I P P M I K N P L C T A E R G C G Q Q I P G L I N N  
1301 ATGCCAAGTACAAAGTCCAGGCGAACTCAATGAAATCATAAATCCCAATTACATGGGGGAATGGCATCCACCGGAAATGAAAACCCGCTATACTACGA  
349 A K Y K G P G E L N E I I N P N Y M G E W H P P E I E N P L Y Y E  
1401 AGAGCAGCACCCATTGCGCATCGAAAACGTTATCAGTGGTGTGATCCTCGAGTTTGGAGTGGATCTCCAAACATGTTGATAAGCAACATTTATGTTGGT  
382 E Q H P L R I E N V I S G V I L E F W S G S P N M L I S N I Y V G  
1501 AAAAATGTAACAGAGGCGCAAATTTGGGAATAAGACTTGGCTGATGAGAGACCGCGCTTTAGAGGCTCCGATGGCCCCACAGAACGCAAATTCATGA  
416 K N V T E A Q I I G N K T W L M R D R A F R G S D G P T E R K F M N  
1601 ATAGCAGACTAGGAAATCTACAAACAATTTCCATAACGAAAGAGAATCCCTAATCCATTTGACCGCATTATAGATCGCATATTAGAGCAACCTCTGAA  
449 S R L G N L Q T T F H N E R E S P N P F D R I I D R I L E Q P L K  
1701 ATTTGTGCTTACTGCGGCCGTCGTGCTCTTGACGACGTCGGTTCTTTGTTGTAGTATTACATAGTGGACAAGTGTAGTTTATAACATGGTCTCAA  
482 F V L T A A V V L L T T S V L C C V V F T \*  
1801 AATTCACCACAACGGCTTCTCTTTTATAGATGGTTAACATTATAGTATCAATATTATCATCATGATTAATGATGATGATAACTTACCCGATGTTA  
1901 AATCTTATTTTTTCATGCAGTAAGTAATCATGCAACAGAAAAACCCGTAATTAAGCGAACATAGAACAACCTAGCATCCCGATAAGACGGATAGAATA  
2001 GTAAAGATTGTGATTTCATTGGCAGGTCCATT

Figure 1. Nucleotide sequence and deduced amino acid sequence of *CNE1* of *S. cerevisiae*. The predicted amino acid sequence is shown in single-letter code. One TATA box and potential polyadenylation sites are underlined.

```

                                1                               38
C      megkwl1cml1vlgttivqaheghdddmiieddllddviee

                                39                               98
H      .ll.vp.llgl.g.vaepavyfkeqf.dg.gwtsrwieskhksdfgk-fvl.-----
S      MKFSAYLWWLFLNLALVKGTSLLSNVTLAEDSFWEHFQAYTNTKHLNQEWITSEAVNNEG
C      ieedvskskppts.ptspkvtykapvpsgevy---.adsfdrgt.sg-.l.k.kkddt

                                99                               158
H      -----sgkfygdeek-----..lqtsqdarfyalsasf.-.fs-nkgq....ftv
O      fgdavr-----..lkttdakfys..akfdksfs-nkgks..i.fsv
S      SK---IYGAQWRLSQGRLQGSAWDKGIVRTGNAAAMIGHLLETPIINVSETDTLTVQYEI
C      ddeiak.dgk.evdemketklpg...lvlmsrakhha.sak.nk.fl-fd.kp.i....v

                                159                               218
H      .heqnid...gyv..fp--nsl-d-qtdmhg.s.-ynim....i.g.gtkk.hvif.y.gk
O      .heqdid...gyv...a--sd.n-.edshge.p-yhim....i.g.gtkk.hvifhy.dr
S      KLDNSLTCGGAFIKLMVSGFMNVEALKHYAPDTEGVELVFGPDYCAPEINGVQFAIN-KVD
C      nfg.gie....yv..l.ktpeln-.dqfhdk.p-ytim....k.gedy-klh.ifrh.np

                                219                               278
H      nvlinkdi.ck-----..deft.....vr-pdntyevk..nsq.esgsled.---
O      nhmikkdi.ck-----..dvft.....vn-.dntyevq...ekaesg.lea.---
S      KITHESKRLRY-LQEMPLSKLTDTSQSHLYTLIIDESAQSFQILIDGKTVMVREHIEDKKK
C      .tgvyeekhakrpdad.ktyftdkkt.....ln-pdn..e..v.qsi.nsgnlln.---

                                279                               338
H      --wd--fl..kk.k.pdas..e...e.ak.d..tds.pe.w-----
O      --wd--fl..kk.k.pdak..e...e.ef.d.eddk.pe.w-----
S      VNFEPPIPTPLMIPDVSVAKPHDWDRIIPDPEAVKLSDRDERDPLMIPHPDGTEPEEW
C      --mt..vn.sre.e.pedq..e...e.pk....d...pd.wn.da.ak..deea.k.dg.

                                339                               398
H      --dk..h.p..d.k..ed.d.emd...e..-----v.q.pe...ew
O      --dk..h.p..d.k..ed.d.emd...e..-----mvd.pe...ew
S      NSSIPEYILDPNQKPSWWKLEHGEWIPPMIKNPLCTAERGCGQIPLINNAKYKGGP
C      ldde...vp..d.e..ed.d.dmd...ea.q.a..k.esap...vwqrpm.d.pn...kw

                                399                               458
H      kprq.d..d.k.t.ih...d..e.spdpsiyayd.-fgvlg.dl.qvksgtifd.flitn
O      kpkqkk..a.k.k.ih...i.d.tpdnlyvydd-.gaigfdl.qvksgtifddvi.td
S      ELNEIINPNYMGWEHPPEIENPLYEYEQHPLRIENVISGVILEFWSGSPNMLISNIYVGK
C      kppm.d...q.i.k.rk.p..dff.dle.fkmtl-f.aig..l..mtsdiffd.fi.cg

                                459                               518
H      deay.eef..e..gvtka.e-----kqmkdkqdeeqrl-----
O      s.e..kkf.e..lkitrege-----k.-kgkktkkqk-----
S      NVTEAQIIGNKTWLMRDRAFRGSDGPTERKFMNSRLGNLQTFHNERESPFPDRIIDRI
C      drrrvddwa.dg.gikka.dgaaepgvvgmieaaeerpwlwvvyvltval.vflv.lfc

                                519                               578
H      -----keeedk-----krkeeeeaedkeddedkdedeed
O      -----kkeeknek-----ikkekmmkrkrankkkk*
S      LEQPLKFLVLTAAVLLTTSVLCCVVFVFT*
C      csgkkqs spveykktdapqpdkveeeekkeekdkgeeegeekleekqksdaeedggt

                                579                               609
H      keedeedvpgqakdel*
C      asqeaddrkpkaeedeilnrspnrkrpre*
    
```

Figure 2. Comparison of the amino acid sequence of *CNE1* (S) with calreticulin from *H. sapiens* (H) and *O. volvulus* (O) and calnexin from *C. domesticus* (C). Identity with *CNE1* is indicated by '.' and gaps are indicated by '-'.

## ACKNOWLEDGEMENTS

We are grateful to P. Linder for helpful comments and to R. Doelz for maintenance of computer facilities. This work was supported by the Swiss National Science Foundation, Grant 31-27923.89.

## REFERENCES

- Devereux, J., Haeblerli, P. and Smithies, O. (1984). A comparative set of sequences analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387–395.
- Fliegel, L., Burns, K., MacLennan, D. H., Reithmeir, F. A. F. and Michalak, M. (1989). Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum. *J. Biol. Chem.* **264**, 21522–21528.
- Kaback, D. B., De Steensma, H. Y. and De Jonge, P. (1989). Enhanced meiotic recombination on the smallest chromosome of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **86**, 3694–3698.
- Murthy, K. K., Banville, D., Srikant, C. B., Carrier, F., Bell, A., Holmes, C. and Patel, Y. C. (1990). Structural homology between the rat calreticulin gene product and the *Onchocerca volvulus* antigen Ral-1. *Nucleic Acids Res.* **18**, 4933.
- Pelham, H. R. B. (1990). The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem. Sci.* **15**, 483–486.
- Rokeach, L. A., Haselby, J. A., Meilhof, J. F., Smeenk, R. J., Unnasch, T. R., Green, B. M. and Hoch, S. O. (1991). Characterization of the autoantigen calreticulin. *J. Immunol.* **147**, 3031–3039.
- Rose, M. D., Novic, P., Thomas, J. H., Botstein, D. and Fink, G. R. (1987). A *Saccharomyces cerevisiae* genomic plasmid bank based on a centromere-containing shuttle vector. *Gene* **60**, 237–243.
- Smith, M. J. and Koch, G. L. E. (1989). Multiple zones in the sequence of calreticulin (CRP55, calregulin, HACBP), a major calcium binding ER/SR protein. *EMBO J.* **8**, 3581–3586.
- von Heijne, G. (1990). The signal peptide. *J. Membrane Biol.* **115**, 195–201.
- Wada, I., Rindress, D., Cameron, P. H., Ouk, W.-J., Doherty II, J. J., Louvard, D., Bell, A. W., Dignard, D., Thomas, D. Y. and Bergeron, J. J. M. (1991). SSR $\alpha$  and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. *J. Biol. Chem.* **266**, 19599–19610.