evident already that not every infant with classical Menkes disease will have the same level of benefit from early copper histidine treatment. While parenteral copper replacement should still be offered to all Menkes disease infants identified at an early age, there is a need to think about additional therapies for this disorder, in my opinion. Improved understanding of the copper transport process mediated by the Menkes gene product and its homologs may eventually facilitate this.

Stephen G. Kaler
Clinical Neuroscience Branch, National Institute of Neurological Disorders and Stroke National Institutes of Health, Bethesda, Maryland and Children's Hospital, Washington, DC, USA

1. Tumer, Z. *et al. Nature Genet.* **12**, 11-13 (1996).
2. Kaler, S.G. *et al. Ann. Neurol.* **38**, 921-928 (1995).
3. Kale,r S.G. *et al. Biochem. Mol. Med.* **57**, 37–46 (1996).
4. Petrukhin, K. *et al. Hum. Mol. Genet.* **3**, 1647-1656 (1994).
5. Kozak M. An analysis of 5' non-coding sequences from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* **15**, 8125-8148 (1987).
6. Grant C.M., Miller P.F & Hinnebusch A. *Mol. Cell. Biol.* **14**, 2616-2628 (1995).
7. Peabody D.S. & Berg P. *Mol. Cell. Biol.* **6**, 2695-2703 (1986).
8. Winnard, A.V., Mendell, J.R., Prior, T.W., Florence, J. & Burghes, A.H.M. *Am. J. Hum. Genet.* **56**, 158-166 (1995).

Internal repeats in the BRCA2 protein sequence

Sir — The breast cancer susceptibility genes *BRCA1* and *BRCA2* are major contributers to a highly penetrant, autosomal dominant predisposition to the disease[1]. Apart from a RING finger identified in BRCA1 (ref. 2), no homology to any other protein in public sequence databases has been found for either of the genes. A granin-binding site has, however, been claimed to be present in BRCA1 (central region) and BRCA2 (C-terminal part) and granin-like properties have been demonstrated for BRCA1 (ref. 3). In addition, a weak similarity between BRCA1 and BRCA2 has been proposed in their central regions[1]. In an attempt to check the significance of this similarity, we studied the respective regions in detail. We were unable to confirm this similarity, but surprisingly, we found eight repetitive units in the BRCA2 protein sequence (Fig. 1).

The complete BRCA2 protein sequence contains 3,418 residues[4]. The eight internal repeats (referred to as BRC repeats) are located in the central portion of the protein and cover nearly a third of the protein. Repeats in BRCA2 have been independently noted (K. Hoffman and P. Bucher, personal communication). All repeats are encoded by the large exon 11 (ref. 4 and R. Wooster, EMBL accession number X95161). As the region of detectable similarity in the core of each repeat is much smaller than the spacing between the repetitive units, domain boundaries are difficult to determine. We thus performed motif and profile searches (for details of the methods see ref. 5) with the core of each repetitive unit in order to identify additional repeats in other proteins. Different methods independently identified a domain in a putative uncharacterized gene (T07E3_2) in *C. elegans*[6] that is significantly similar to the repeats in BRCA2 (Fig.1). The N-terminal location of the identified domain in the *C. elegans* protein allows the phasing of the repeats in BRCA2 to be defined and a first rough assignment of the domain borders to be made. The length of a minimal BRC domain should not exceed 80 amino acids, as this is the length of the seventh repeat in BRCA2 which seems to be the shortest one. At least four sec-

```
2D structure:                              eEEe   eeeEEe HHHHHHHHHHHHh                                  eeeeeee
BRC1        987 NDYMNKWAGL.LGPISNHSFGGSFRTASNKEIKLSEHNIKKSKMFFKDIEE.....................QYPTSLACVEIVNTLAL.DNQKKLSKPQSINTV
BRC2       1198 NDCNKSASGY.LTDENEVGFR.GFYSAHGTKLNVSTEALQKAVKLFSDIENISEETSAEVHPISLSSSKCHDSVVSMFKIENH..NDKTVSEKNNKCQL
BRC3       1407 EQLTATKTE..QNIKDFETSDTFFQTASGKNISVAKESFNKIVNFFDQKPEEL....HNFSLNSELHSDIRKNKMDILSYEET..DIVKHKILKESVPV
BRC4       1501 NQLVTFQGQPERDEKIKEPTLLGFHTASGKKVKIAKESLDKVKNLFDEKEQGTSEIT.............SFSHQWAKTLKYREACKDLELACETIEITAA
BRC5       1649 KSPATCYTNQ.SPYSVIENSALAFYTSCSRKTSVSQTSLLEAKKWLREGIFDGQPER.............INTADYVGNYLYENN..SNSTIAENDKNHLS
BRC6       1822 NKNAAIKLSI.SNSNNFEVGPPAFRIASGKIVCVSHETIKKVKDIFTDSFSKVIKEN......NENKSKICQTKIMAGCYEALDDSEDILHNSLDNDEC
BRC7       1955 SDICKCSIGKLHKSVSSANTCGIFSTASGKSVQVSDASLQNARQVFSEIEDS.....................TKQVFSKVLFKSNEHSDQLTREENTAIRT
BRC8       2036 PEHLISQKGFSYNVVNSSAFS.GFSTASGKQVSILESSLHKVKGVLEEFD.....................LIRTEHSLHYSPT.SRQNVSKILPRVDKR
CET07E3_2    12 FDTISEPDSFDEPKGVPISMEPVFSTAAGIRIDVKQESIDKSKKMLNSDLKSKSSSKGGFSSPLVRKNNGSSAFVSPFRREGT.SSTTTKRPASGGFED
consensus       tt.ht...t...t.....t...hF.TASGKthtVtttShtK K hhtt.t.....................thht.h.htt...t.thht...t....
```

Fig.1. Multiple alignment of the 8 repeated domains in BRCA2 (EMBL accession number U43746) and the N teminus of the *C. elegans* CET07E3_2 sequence[6] (accession no. U13643). First column: names; second column: position of the displayed regions in the respective sequences. The secondary (2D) structure elements were predicted using the PhD server[8]: H/h denotes a helix and E/e a β strand with an expected accuracy higher than 82% (upper case)/72%(lower case). Amino acids conserved in at least two-thirds of the sequences are highlighted; hydrophobic residues conserved in all but one sequence are given in bold. The consensus line summarizes residue properties conserved in at least 80% of all sequences (t — turn-like or polar, h — hydrophobic). The internal repeats in BRCA2 can be detected by independent approaches such as i) BLASTP searches[9] with arbitrarily chosen fragments of BRCA2 that indicate internal hits with a similar matching behaviour, ii) dotplot analysis that reveales suboptimal hits parallel to the diagonal, and iii) the REPEATS program (M. Vingron, Heidelberg, unpublished) that gave significant results (more than 6 standard deviations above the mean calculated for random sequences). With the core of the BRC repeats in hand, several motif and profile search methods identify the *C. elegans* protein. For example, the motif search program MoST[10] using the central conserved block in the alignment identified the *C. elegans* protein with an r value (the ratio of observed versus expected hits) of 0.0045 whereas all other proteins in the databases had r values larger than 0.1. Profile searches[11] with an alignment of the 8 repeats in BRCA2 as input revealed the *C. elegans* protein as the best hit with Z scores of Z = 5.91 and total score t = 11.14 whereas other proteins scored below Z = 5.18 and t = 9.37.

ondary structure elements are present in a single repeat (Fig.1) pointing to a globular domain. This is supported by the single occurrence of a BRC domain in the *C. elegans* protein. The segments between the repetitive BRC domains are variable in length and predicted to be mainly non-globular segments that might have spacer functions or that might carry exposed binding regions.

The segment of proposed homology[1] between BRCA1 and BRCA2 overlaps the portion of BRCA2 covered by the BRC repeats, but the conserved residues are not consistent with the conservation pattern of the BRC repeats. Thus, we conclude that this similarity is most likely spurious. For the same reason, a recently proposed homology to yeast ORF YER033c[7] becomes insignificant.

Although our findings do not give any additional clues as to the function of BRCA2, the identification of internal repeats reveals the architecture of large parts of this modular protein. The probable globular nature of the BCR repeats and the occurrence as a single unit in another protein should allow the experimental analysis of individual BRC domains. We are thus attempting to solve the three-dimensional structure of some of the BRC domains to gain further information on possible ligand-binding features.

Peer Bork[1,2]
Niklas Blomberg[1]
Michael Nilges[1]
[1]*EMBL, Meyerhofstr. 1, 69012 Heidelberg, Germany*
[2]*Max-Delbrück-Center for Molecular Medicine, 13122 Berlin-Buch, Germany*

1. Wooster, R. *et al.* Nature **378**, 789–792 (1995).
2. Miki, Y. et al. Science **266**, 66-71 (1994).
3. Jensen, R.A. *et al. Nature Genet.* **12**, 303–308 (1996).
4. Tavtigian, S.V. *et al. Nature Genet.* **12**, 333–337 (1996).
5. Bork, P. & Gibson, T. *Meth. Enzymol.* **266**, 162–184 (1996).
6. Wilson, R. *et al. Nature* **368**, 32–38 (1994).
7. Oliver, S.G. *Nature* **379**, 597–600 (1996).
8. Rost, B. *et al. CABIOS* **10**, 53–60 (1994).
9. Altschul, S.F. *et al. Nature Genet.* **6**, 119–129 (1994).
10. Tatusov, R.L. *et al. Proc. Natl. Acad. Sci. U.S.A.* **90**, 12091–12095 (1994).
11. Gribskov, M. *et al. Proc. Natl. Acad. Sci. U.S.A.* **84**, 4355–4358 (1987).